

GenCore version 5.1.6
Copyright (C) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: July 30, 2004, 14:16:25 ; Search time 3640 Seconds
(without alignments)
7251.623 Million cell updates/sec

Title: US-09-868-338-7_COPY_1117_1725
Perfect score: 609
Sequence: 1 atgataagaatcaatgacct.....cggataaccattatcaaaacta 609

Scoring table: OLIGO NUC
Gapop_60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size : 0
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl: *
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.on.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.tod.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	21	3.4	110000	8	AC145127_20	Continuation (21 o
2	21	3.4	153510	8	AC099734	AC099734 Oryza sat
3	21	3.4	303511	8	AE017053	AE017053 Oryza sat
4	21	3.4	313050	1	BX321857	BX321857 Nitrosomo
5	20	3.3	2080	10	MMU011080	AJ011080 Mus muscu
6	20	3.3	22190	9	AC117443	AC117443 Homo sapi
7	20	3.3	86385	2	AC138385	AC138385 Mus muscu
8	20	3.3	86385	2	AC138385	AC138385 Mus muscu
9	20	3.3	121201	9	AC128685	AC128685 Homo sapi
10	20	3.3	154934	2	AC090177	AC090177 Homo sapi
11	20	3.3	157793	9	AC090181	AC090181 Homo sapi
12	20	3.3	171061	2	AC109266	AC109266 Mus muscu
13	20	3.3	175066	9	AC005900	AC005900 Homo sapi
14	20	3.3	180172	2	AC115834	AC115834 Mus muscu
15	20	3.3	184180	9	HS1042K10	AL022238 Human DNA
16	20	3.3	221618	2	AC098411	AC098411 Rattus no
17	19	3.1	315	10	MUSZEP92	D10631 Mus musculu
18	19	3.1	855	6	AR317790	AR317790 Sequence
19	19	3.1	2202	9	BC034912	BC034912 Homo sapi
20	19	3.1	3582	10	BC053927	BC053927 Mus muscu
21	19	3.1	32436	1	AB101202	AB101202 Acinetoba
22	19	3.1	45643	3	U23177	U23177 Caenorhabdi
23	19	3.1	64201	2	AC117833	AC117833 Homo sapi
24	19	3.1	64900	2	AC121510	AC121510 Mus muscu
25	19	3.1	67339	2	AC138375	AC138375 Mus muscu
26	19	3.1	68930	9	AC068136	AC068136 Homo sapi
27	19	3.1	82165	9	AC117471	AC117471 Homo sapi
28	19	3.1	85397	3	AC026301	AC026301 Caenorhab
29	19	3.1	90024	2	AC139690	AC139690 Homo sapi
30	19	3.1	97137	9	AC004745	AC004745 Homo sapi
31	19	3.1	110000	9	AE014305_0	AE014305 Homo sapi
32	19	3.1	118401	9	AC091305	AC091305 Homo sapi
33	19	3.1	129245	9	AL359547	AL359547 Human DNA
34	19	3.1	133652	9	AC010326	AC010326 Homo sapi
35	19	3.1	142624	2	AC128628	AC128628 Rattus no
36	19	3.1	152631	9	AC117432	AC117432 Homo sapi
37	19	3.1	155656	9	BX293560	BX293560 Human DNA
38	19	3.1	155757	2	AC079984	AC079984 Homo sapi
39	19	3.1	156678	2	AC145026	AC145026 Sus scrof
40	19	3.1	158770	9	AC007321	AC007321 Homo sapi
41	19	3.1	162475	9	AC146163	AC146163 Pan trogl
42	19	3.1	163288	2	AC119264	AC119264 Mus muscu
43	19	3.1	164061	9	AL445188	AL445188 Human DNA
44	19	3.1	164394	9	AC116904	AC116904 Homo sapi
45	19	3.1	167398	9	CNS072ET	AL445886 Human chr

ALIGNMENTS

RESULT 1
AC145127_20
WPCOMMENT
Sequence split into 24 fragments LOCUS AC145127 Accession AC145127

Fragment Name	Begin	End
AC145127_00	1	110000
AC145127_01	100001	210000
AC145127_02	200001	310000
AC145127_03	300001	410000
AC145127_04	400001	510000
AC145127_05	500001	610000
AC145127_06	600001	710000
AC145127_07	700001	810000
AC145127_08	800001	910000
AC145127_09	900001	1010000

/note="Similar to Oryza sativa gene, repeat sequence
Micropan-1"
5518..5539
/note="Simple Sequence Repeat (AT)n"
5542..5575
/note="Simple Sequence Repeat (AT)n"
5602..5627
/note="Simple Sequence Repeat (AT)n"
5724..5745
/note="Simple Sequence Repeat (AT)n"
568..6457
/note="Simple Sequence Repeat (TA)n"
8249..15154
/gene="OSJNBb0004A06.3"
/note="Unknown protein"
join(8249..8785,9076..9120,9930..10065,11299..11387,
11908..11950,12602..12789,13220..13423,14642..15154)
/gene="OSJNBb0004A06.3"
/codon_start=1
/product="Unknown protein"
/protein_id="AA111187.1"
/db_xref="GI:22953954"
/translation="MNLNNGKGLTDHNMKKQKKEKGRSDWHGSGTKDTIPFM
GDADNADGVDVAGCTGMAFALLLCPHOLMOIQKCCVYTSSASGSGR
GQAWAMVIMRPADEGTADEDEGALLIAROGDRQKGRWRPRAERQ
APFYELASRRRDGSPVTRSSYSLITGQALNQRYKVAWTERSLIGS
SLAQSPSCGRNRNGSGIKFILKOLFVDESHLEFLNHRDLPVCEAFKRSK
TKQNLQSSIKCNITWPGGSLHLLQFQDVFIRDFRGNKYLIVILRRSG
IICGAKETPKPIFKHRSFRPVLRIENGDSMTTICQGVNSGIIKTPFH
ICKSVQFQNIILLIGSPSEAFLEMEENKNSIVKFRHFSPQNFQEQ
QKQSGSGTQSLMPLDKKAKKALTLAGCTVQGPSPQAVHTTASCKEKKDK
EPNQSTPEVGTREMLVHSGTAKKLVSSITIEDHVSCKSNVNTVGVKHMPEIQSPV
LVRSVSPSASHQNRFT"
9056..9073
/note="Simple Sequence Repeat (AT)n"
10560..10589
/note="Simple Sequence Repeat (AT)n"
10975..11020
/note="Simple Sequence Repeat (AT)n"
complement(19177..22551)
/gene="OSJNBb0004A06.4"
/note="Contains similarity to disease resistance protein"
complement(join(19377..21480,22331..22551))
/codon_start=1
/product="Putative disease resistance protein"
/protein_id="AA111188.1"
/db_xref="GI:22953955"
/translation="MGRGTEGRSGAGSGHPAPVAGKEREVEASGLPTLSLVE
EVEKEVNGRRRGKVRGAKRQMGSGGRYILDDVDSAKTKWTLISYLLDNKE
RSRLIVSRQAVGCTCPENKDLPIPSLRDSKELFNRSSEKSTDRKHVQ
NNVFDLWRCGGQPLAVTMAGLVACNQQPKSYWAGLKLPEEVSITAGAQEQT
SLNDGVTRILDCNNLPGYRLTCLLYLAIFPKGKISRKCLSRWIAEGFVNAQQT
LTABEAYSFNLHRLKLRIPVHSSNGKLTFOVDMVLDIVTKAREENFTVQV
GLHMAIPNNKVRSLQSSSSKHGSDTKGNLSQVSLTVPFSLTPEAFNDRI
LOVDFQGLGKFKHMKICKMFLVXLSLCTDITHVPTIKLEYLTDIRETR
VKELPKEVEQLKLSRILGSKSNKPKRGLRQPKSKQKHSMLTQKKEGMKAL
RILSGIKDITVAGHUTGLKLTLYKLNPEEPGRKILTERSSIEYLSGCG
LQTAINEGSGNFINLGNMSPRYLVALELSGMLKPPGKWLRTLSLKLITSL
VLRTDLEHALPLSLTFSFGSEMEKQDKMKDIIEDNKSLSGDSIFVPGEGFK
SLKLRLFPVLPKLFCHNAPALEMIQFOAFGLFGLDITLGNLKGVRLEAKPR
BEDQTAQINDLLVRLDKDSTEGKVIIDHTFTS"
complement(24383..25405)
/gene="OSJNBb0004A06.5"
/note="Contains similarity to disease resistance protein"
complement(24383..25405)
/gene="OSJNBb0004A06.5"
/codon_start=1
/product="Putative disease resistance protein"
/protein_id="AA111189.1"
/db_xref="GI:22953956"
/translation="MELVAGSTATWDSLLGKLNLLACEYDILIRVGRDIOYISDEL
ASMAKAFLLDLAREDPNRKKNWKQIRDMAYDCEDCIDDFAHRLPNDLSLAKCCPWV

TLVYDLWTFGRPREIASSIAELKVRALQIADRIRYGVNPNVTQKKGPPDATSDVIA
EDQASHLGNFPGVMEKAKDLREWVGTAQCEFAVVGIVGFGKTTAMALYK
KMVQFDCRAWTVQNYDLVDLNDILKQIDYQCCSKTGTSENITKLARFGSK
LKEDVQRTGSLROSSPRSBETSNLKRITETDNKLESQIKLLDKRCVHPHSILSIS
ISISY"
repeat_region 26913..27062
/note="Similar to Tourist Oll MITE element"
complement(27190..27338)
repeat_region
Query Match 3.4%; Score 21; DB 8; Length 153510;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 340 CAAGTACTTTCATGCTGTGGT 360
|||||
DB 110715 CAAGTACTTTCATGCTGTGGT 110735
|||||
RESULT 3
AE017053 303511 bp DNA linear PLN 06-JUN-2003
LOCUS
DEFINITION Oryza sativa (japonica cultivar-group) chromosome 10, section 7 of
77 of the complete sequence.
ACCESSION AE017053 AE016959
VERSION AE017053.1 GI:31430010
KEYWORDS
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE
1 (bases 1 to 303511)
AUTHORS
CONSTRM The Rice Chromosome 10 Sequencing Consortium
TITLE In-depth view of structure, activity, and evolution of rice
chromosome 10
JOURNAL Science 300, 1566-1569 (2003)
REFERENCE 2 (bases 1 to 303511)
AUTHORS Buell, C.R., Wang, R.A., McCombie, W.R., Messing, J. and Yuan, Q.
TITLE Direct Submission
JOURNAL Submitted (05-MAY-2003) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
COMMENT This is the pseudomolecule for rice chromosome 10, which was
constructed by resolving discrepancies between overlapping BACs,
trimming the overlap regions, and linking the unique sequences to
form a contiguous sequence. Genes in individual BAC clone were
identified by a combination of several methods: Gene prediction
programs, searches of the complete sequence against a peptide
database and EST databases. Genes with similarity to other proteins
are named after the database hits. Genes without significant
peptide similarity but with EST similarity are named as unknown
proteins. Genes without protein or EST similarity, that are
predicted by more than two gene prediction programs over most of
their length are annotated as hypothetical proteins. Genes
encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy,
http://genome.wustl.edu/eddy/tRNAscan-SE/).
FEATURES
location/Qualifiers
1..303511
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="10"
misc_feature
1..79862
/note="Chromosome Sequence Derivation: nucleotide sequence
in this region was derived from BAC clone OSJNBb0072F04
(GB:AC092553)."
complement(<6121..>7701)
/locus_tag="OSJNBb0072F04.10"
/note="Contains similarity to NPK1 related protein kinase"
complement(<6121..>7701)
/locus_tag="OSJNBb0072F04.10"
complement(6121..7701)
gene
mRNA
CDS

```

/locus tag="OSJNBb0072F04.10"
/codon_start=1
/product="putative NPX1-related protein kinase"
/protein_id="AAP51982.1"
/db_xref="GI:31430011"
/translation="VAAAIIGGGGWTLRISVGRGASGAVLSLAANDVSGELFVKSAG
EGARQQLREWSVMSGLSPHLVCLGFGVQAGCGGGEQLEFVAPGSLADVA
RNGRLDEGAVTYADVLRLGLDHLGVLVGHGDKGNSLVNVLGADGAKLDTFCARV
ANFGGSKQVLGGTFAFPAPEVARGEEOGLAADVWALCQIVEMATGRAPWSDNVL
PALHKIGYDADLPDLRWLSPEAKDFLRCQLQRAGDRPTAAQLLOHPPTSKSCGLN
KTKVATWSPSALDAILWSESSSTDEEDVDMSSNPTGRIRAMACSCQTLPDWD
SDHGCSWLEVLGSVINVAKTARIVTSMACSPSSVPDMSQCGWIDLVLSVS
ISIAKLETRATRADVSSECPAKVRMACSPSSVPDMSQCGWIDLVLSVS
EDVAAAOQISQAGVSGISLVGVSGSEQSVVVENQEDFTSLSSCSERVLLVGHAA
NNAASKKAGIKCSNFSC"
complement (<10528..>12105)
/locus tag="OSJNBb0072F04.11"
/notes="contains similarity to NPX1 related protein kinase"
complement (<10528..>12105)
/locus tag="OSJNBb0072F04.11"
complement (<10528..>12105)
/locus tag="OSJNBb0072F04.11"
/codon_start=1
/product="putative NPX1-related protein kinase"
/protein_id="AAP51983.1"
/db_xref="GI:31430012"
/translation="VAAATIGGGGWTLRISIGHGASGATVLSLAADASGELFVVKV
ADVAATARQQLREWSVMSGLSPHLVCLGFGVQAGGAGGSEQLLLEFVAGGSLA
DVARNGRDLSSAFRAYADVRLGLDYLHLEKLVVGHGVKGSNVLVGADGAKLADFG
CARVWFGSKQVLGGTFAFPAPEVARGEEOGLAADVWALCQIVEMATGRAPWSDM
DVLAAIRMGYDADLPDLRWLSPEAKDFLRCQLQRAGDRPTAAQLLOHPPTSKSC
GLKEVYKATWSPSALDAILWSESSSTDEEDVDMSSNPTGRIRAMACSCQTLPDWD
TLPDMSDHLGVLGVTGINSIAKKTAAAEVLSGSPSSVPDMSQCGWIDLVLSVS
SDNHGIDVLSPADDDNGEQAPEEDVVAADQIFGEAVGSIIVGVSGSEQSVVVE
NORDEFISLVCYSCEIRLIVAVHADNNAASKKAGIKESCHDPRPSIRCAHNLFLSL
IFEQIDNSNVPSAKY"
complement (<16098..>17177)
/locus tag="OSJNBb0072F04.12"
/notes="unknown protein"
complement (<16098..>17177)
/locus tag="OSJNBb0072F04.12"
complement (<16098..>17177)
/locus tag="OSJNBb0072F04.12"
/codon_start=1
/product="unknown protein"
/protein_id="AAP51984.1"
/db_xref="GI:31430013"
/translation="WAKPRTSLPRLVIGARWDAERPLGRLLILAHAAFLDAGFVPA
AARADNSIRLPRKVGRTASSLPRLYAAPQLLHWPDDAAVQLRLCAHGRHLVYVSN
RCSMFEMLDVTLVCLDAAALGGLADDTARALRDARLAALWALADRLCRRVL
VDVCAENGVTLPPTFMSLPDDVKAAILARLPDGDLDLAPAECTCAGLRRLVADRDRD
LANKPRYKLPFLQLIGSDDDDDGPPTTEVSKKKYVAARLWLPGEIIFASMRTRPL
IYAPLLDLDFTSEFTFWPDDKPSLPPEIIVPERRRRRRWRAMPDAGHGLAARGH
GDKPRHAGAGVHSPSRFRWKR"
complement (<19076..>20170)
/locus tag="OSJNBb0072F04.13"
/notes="unknown protein"
complement (<19076..>20170)
/locus tag="OSJNBb0072F04.13"
complement (<19076..>20170)
/locus tag="OSJNBb0072F04.13"
/codon_start=1
/product="unknown protein"
/protein_id="AAP51985.1"
/db_xref="GI:31430014"
/translation="WADTSLPHRIITCAAMDAEPLGLRLVILAHAAFLDAGFVSTGAA
NIDGACQSVLRQVLSASLSLRYAPQLLHROPDAAALAAAAAVALRVCAGRRH
VYVYCVRFANPLDTYICIDAPAAALLAGGLDDTARALARERCARLAAWSLA
DLRCRLVYDLCAKNGVPEPEHLEMSLPDDVKVAILARLAAGARVCTVCGNL
LVAHDSITLWKMVYKLSQRLRLRLFLVGSYGETAVSKAVVAVRERVPAADV
FWGEIILPFTWEMRVIRIYIPFPFPFPFPFPFPFPFPFPFPFPFPFPFPFPFP
AAPGHGDKQWRGAGVHSPSSGFRWKR"
complement (<25322..>29537)

```

```

/locus tag="OSJNBb0072F04.14"
/notes="contains similarity to copia type pol polyprotein"
complement (<25322..>26796,27148..>27678,27734..>27814,
28019..>28279,28375..>28542,29526..>29537))
/locus tag="OSJNBb0072F04.14"
complement (<25322..>26796,27148..>27678,27724..>27814,
28019..>28279,28375..>28542,29526..>29537))
/locus tag="OSJNBb0072F04.14"
/codon_start=1
/product="putative copia-type pol polyprotein"
/protein_id="AAP51986.1"
/db_xref="GI:31430015"
/translation="MSYKALVAKENVMIVDSCVSRHMTGDKNWFSLKASKTSRII
FEDASTSVLAIGLVKVKCKTKSKVDFSCDSDVLIIISYGRVFXADFPNVPVITC
LAAPFDKDMFWRRLRGVDFDLRLSLGDLVGRPLKKNLIDLEAFCHERGLFLR
LDLEFPGLSKRIFSDNGLSHEFSSRPQONGVVERKVLVEMARMMDLDEKTFPK
FWAEINTACYISNRFLSKLGTLETKIETCEVDFEASPTRIDIASTISQVOGE
EARTDGLGYPHARGYRLIETNKIETCEVDFEASPTRIDIASTISQVOGE
DGRIFDESDDDNNDVGSALFNERNKVTVEPPEHNFPGTKWVQMDVQSDGL
IVRNKTRVAQGFQVEGLDDETFARVARVEAIRLLAFATSKGFKLVQMDVQSDGL
NGPTQBEVYKQPPENPEPNHVKLSKALYKLAQPRAMVDREKNFLAKGSMG
KDKILFVLKXGNQLFVQIYVDLIFGSGTHALVVEPAETWREREMSGELSYEL
GLQIKTPQGTFFVHMKITKDLRRFMENCKPISPIGTAVLDDEDECAVDQKEY
RSMIGSLYLIAERLDIQLAMCLYTRFPASPRASHQAIRKIMYLNHLEFEGIWST
SSISLGSYDTPFGCRIDRKTSGICHPLETSLIASSRKQSSVAAQTSSESVAA
ASCCSQDLWLSTLKYSLTFEKVPLFCNTSAINAKPNVORSHRTKINICFHLRD
HVEKGDELQFLDKMLQLADIENKPLDSNHFALRGELGIHPFGVW"
complement (<32428..>42847)
/locus tag="OSJNBb0072F04.15"
/notes="contains similarity to transposase"
join (<32428..>32464,33104..>33314,33384..>33512,33627..>33798,
33937..>34750,34784..>35140,36883..>36958,38204..>38231,
39651..>39752,40093..>40296,40509..>40808,41038..>41846,
42146..>42316,42395..>42695,42788..>42847)
/locus tag="OSJNBb0072F04.15"
complement (<32428..>32464,33104..>33314,33384..>33512,33627..>33798,
33937..>34750,34784..>35140,36883..>36958,38204..>38231,
39651..>39752,40093..>40296,40509..>40808,41038..>41846,
42146..>42316,42395..>42695,42788..>42847)
/locus tag="OSJNBb0072F04.15"
/codon_start=1
/product="putative transposase"
/protein_id="AAP51987.1"
/db_xref="GI:31430016"
/translation="MVKDSKTPVYNGWPKQGNIDIVYFKPLVDLHLLTWEGVQWMD
TYKHEFTCAAILGITDYPAIGNTSGOMVCKMCAVFLRDEHPHONKKAQDGTV
EIDRPIFRGSSILAEVKLHERRARGIVEHTWQDEGHFASTERSHWKTSKGCIL
KTEMAATTYLRHAILLVOKRLLEFVLFDDIMVHLVHLVEQVKGILGTAYLDOSMAFRFWM
ASLGMQEDTILNLCQDMLFPLSFDDIMVHLVHLVEQVKGILGTAYLDOSMAFRFWM
GRIKKYNNRSIPEGSIIEGYATEVEVEFIDYMDIQLPIGVPPREHGMKGILGH
RRIIVAPNRQIRIHAFAVLSQALVAPYMDMLDLVLRDNRPRKAALAKMHLQFGK
WMLVYLTADIDGAFSLRSWPNTHMYQYDMNGYTFYTSQDNKSINQNSGVRI
AYDESYESLDYDALVPLFQCKVKVILPGLKIDKHGTIIVDLNDVQDEPFLVAKS
VKQIPYKDSIMKPHVILDKOSIVGMEISTKMTTINPMSHRHPQLMSLKTROA
LAKMFKASAIVKSGSVRRGGGSGNDGPMPSRCVKYEVCTEHEKFASSSEGEHH
ASPTFNLNQHSEEGEBCWRKXLSFTPEGBELAKGALCKMGLFLEFWTDL
NVDFVQCNRTDPFEKMYQKLADYDFEGSTIGSGELVALGHQHELKEITEAISL
ARQGFVREKDEITVALGTEHPGRMGVSPWPKGFPDDTMYRARSPTREE
EAGRMLEVAYQSNITOEIARNVAHPICAGQTIHGVLPADHSRVEVEVPSQRCYV
LOFPLEBQVDTLSAALHNLVWPRIIVITNDGAGDSADSSGLNGRSHSPRSP
SLPFRQTPQPOPSQKSPSTGRKKTTPSASQMSGSSSMERLPHSKSLVNPVPT
SLPKVSGSLTSLGPKKPERKIPKVIHFIKCAQPLPPKPIPSDFRFTLEKSE
MQSRSLFLDKASDEAAEKSDVLKSLDDINDLKDRAVLSNMDDPMRSCYOCRS
LDKRIAFDPAVNVFNQLSKEREIDYFNALVKQNGDHLILLPILSHHLLVILN
IDDSKICVDSLRKGTNYQTIIMNALNRAVYRRKRTYGRKTCADATISFRIFENQYI
YRQPALNLCOMVYMYLMCFVSGHLLPRNAEKIGLSETSEMLPHVFTALTD"
complement (<48333..>52859)
/locus tag="OSJNBb0072F04.16"
/notes="contains similarity to NBS LRR type resistance
protein"
complement (<48333..>48348,48499..>48586,49139..>52859))
/locus tag="OSJNBb0072F04.16"
complement (<48333..>48348,48499..>48586,49139..>52859))

```



```
/locus_tag="OSJNB0072F04.16"
Query Match      3.4%; Score 21; DB 8; Length 303511;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 340 CAAGTACTTCATGCTGGTGGT 360
|||||
DB 218168 CAAGTACTTCATGCTGGTGGT 218188

RESULT 4
LOCUS BX321857/c 313050 bp DNA linear BCT 23-APR-2003
DEFINITION Nitrosomonas europaea ATCC 19718, complete genome, segment 2/10.
ACCESSION BX321857 AL954747
VERSION BX321857.1 GI:30138174
KEYWORDS complete genome.
SOURCE Nitrosomonas europaea ATCC 19718
ORGANISM Nitrosomonas europaea ATCC 19718
Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
Nitrosomonadaceae; Nitrosomonas.
REFERENCE
AUTHORS Chain, P., Lamerdin, J., Larimer, F., Regala, W., Land, M., Hauser, L.,
Hooper, A., Klotz, M., Norton, J., Sayavedra-Soto, L., Arciero, D.,
Hommes, N., Whittaker, M. and Arp, D.
TITLE Complete genome sequence of the ammonia-oxidizing bacterium and
obligate chemolithoautotroph Nitrosomonas europaea
J. Bacteriol. 185 (9), 2759-2773 (2003)
JOURNAL 22586410
MEDLINE 12700255
PUBMED 2 (bases 1 to 313050)
REFERENCE
AUTHORS Larimer, F.
TITLE Direct Submission
JOURNAL Submitted (12-NOV-2002) Submitted on behalf of the Nitrosomonas
genome consortium, the DOE Joint Genome Institute, Production
Genomics Facility, 2800 Mitchell Drive, Walnut Creek, CA 94598,
USA, and the Genome Analysis Group, Oak Ridge National Laboratory,
1060 Commerce Park Drive, Oak Ridge, TN 37831, USA;
larimerf@ornl.gov
REMARK Nitrosomonas genome consortium
FEATURES
source
1. 313050
Location/Qualifiers
/organism="Nitrosomonas europaea ATCC 19718"
/mol_type="genomic DNA"
/strain="ATCC 19718"
/db_xref="taxon:228410"
complement(132..503)
/locus_tag="NE0270"
complement(132..503)
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="CAD84181.1"
/db_xref="GI:30138175"
/transl_table=11
/translation="MPEQITRKASPFCLRLTPERTLLERAAGLPLGEVIRQOVFE
NRVKRSRKQPKVDHRLSLLGELGRSLANNQLARAACGLLNLTPVEKTSLL
NACADIRHRTLMKSLGN"
complement(885..1253)
/locus_tag="NE0271"
complement(885..1253)
/locus_tag="NE0271"
/function="InterPro IPR002559"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="transposase IS4 family"
/protein_id="CAD84182.1"
/db_xref="GI:30138176"
/transl_table=11
/translation="MFLSPQAGDAPEGRKLLKLENKGFSDTHVIMDKAYEGDETR
QLVLDLGMIPVVPKANRVSPEYDVMYKKRNEVERLFRRLKFRFRIFSRFDKLDVS
FRFIHFALIVDHLISVNRP"
complement(887..1698)
/note="ISNE5"
/locus_tag="NE0272"
complement(1253..1660)
/locus_tag="NE0272"
complement(1253..1660)
/locus_tag="NE0272"
/function="COGS COG3293"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="possible transposase"
/protein_id="CAD84183.1"
/db_xref="GI:30138177"
/translation="MEISASQFKLIENLPIQRGNTVLNVLNAILVYAEKGCKWR
GLPVKGNWHSIYTRANRWARGVLDRLALQQNKLIQLEVDHMSLDSTIVAKHPDG
TGALKKTVFKLSVNHGAGLPKFTWSRQMPER"
complement(1853..2488)
/gene="cyc"
/locus_tag="NE0273"
complement(1853..2488)
/gene="cyc"
/locus_tag="NE0273"
/function="InterPro IPR003088 COGS
COG2863"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="Cytochrome c, class I"
/protein_id="CAD84184.1"
/db_xref="GI:30138178"
/translation="MKFPGTVAIALTFSGVFABSPAGDVEKKEIAAGICACGH
NADGNSAIPYPILAGQYPGYIAKQNDPKVEGETVKRDNQIMAPMVAATLQDDIEN
LAAYSQKPKPGTASDASLVETGKLYQGNLNSIPACSSCHSPNGGQIPHPYPRI
DQHPAYTLSQLAQRQGRKNDTNTMTQIVSRMSEMQAVSEYIATLR"
complement(2423..2488)
/gene="cyc"
/locus_tag="NE0273"
/note="Signal predicted by SignalP 2.0 HMM (Signal peptide
probability 1.000) with cleavage site probability 0.999 at
residue 22"
2595..3197
/locus_tag="NE0274"
2595..3197
/locus_tag="NE0274"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="CAD84185.1"
/db_xref="GI:30138179"
/translation="MRKFSVFLLANIFVVFYLHGRPDNLPAQIALIHSEKIELLPA
KVACLKWNLIQVVOHVRVEISKWESGQDHTEISRGVTVHWHVHPIPLRNARETAK
QIEQLKXSGISYLIHQENADSPWENAIISLALPDDSDVAALVELKKGKVERIMDSEQ
VLEQFEFDIRNFTQEVSRQLAQQFPETKLEVTCSRJ"
2595..2660
/locus_tag="NE0274"
/note="Signal predicted by SignalP 2.0 HMM (Signal peptide
probability 0.892) with cleavage site probability 0.890 at
residue 22"
3311..4129
/gene="map"
/locus_tag="NE0275"
3311..4129
/locus_tag="NE0275"
/function="InterPro IPR001714:IPR002467 COGS
COG0024"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="Methionine aminopeptidase"
```

/protein_id="CAD84186.1"
 /db_xref="GI:30138180"
 /translation="MQAGAGVVIHIKTAPEHETMTRIAGRLASEVLDVIEPVVPGVTT
 GNDELCHRYMDVQXTIPAPLVNAPPGYSPYKSICTSVNHWGCHGVGPKLKDGD
 VNLDTITVYEGVHGTSMYVYGEPFSIQAKRLCELTVEAMWRIETIRPKHLGDIG
 HAJORLAEGAGSVVREFCGHIGAKFHEDPOVLHYGRAGTGIELKPCMFTVEPMIN
 AGKAAIKQLPDQWTYITKDHLSAQWEHTILVTDSEYVLTVSSGSPARPAYRF"
 4197..4922
 /gene="rph"
 /locus_tag="NE0276"
 4197..4922
 /gene="rph"
 /locus_tag="NE0276"
 /EC_number="2.7.7.56"
 /function="interPro IPR001247:IPR002381 COGS COG0689"
 /codon_start=1
 /evidence=not_experimental
 /transl_table=11
 /product="3' exoribonuclease family"
 /protein_id="CAD84187.1"
 /db_xref="GI:30138181"
 /translation="MPCNNRAPAQMVRVRIIRHYVHAEGSVLIEYGETRVICTASV
 IEKVPPLKAGGQWLTARYGMLPSTGTCWRQRAKQKQSGRTWELQILGRALRSI
 LDLEKLGEBTIQYDCDQVQADGTRTASITGAFVADYDAIDYLAERMIQNPIDRV
 AVVSGILKQGLDLDLHEDSGCDTDLNVMTGSLGLVEVQGAERKVVFSRQELVM
 LNAQQGLQELFEDVQRKALETVA"
 4960..5571
 /locus_tag="NE0277"
 4960..5571
 /locus_tag="NE0277"
 /function="interPro IPR002637 COGS COG0127"
 /codon_start=1
 /evidence=not_experimental
 /transl_table=11
 /product="Ham1 family"
 /protein_id="CAD84188.1"
 /db_xref="GI:30138182"
 /translation="MNKIVIASNNAGKLAETSRLLAPLGIEVWTSQSLGVTEADPHM
 TFVENAKARHASLATGIPALADSGICVSALRGDGVFSVAGVPRSDERNRKL
 VEALHGSDRRAYVCVIVLPHGDQDQPVLIETWGEIITAEPIGOGGYPHPFL
 PELGKTAELSIEENRSHRGKALRLVQLSENETVPVVPV"
 5608..6807
 /locus_tag="NE0278"
 5608..6807
 /locus_tag="NE0278"
 /function="interPro IPR003401:IPR004559 COGS COG0635"
 /codon_start=1
 /evidence=not_experimental
 /transl_table=11
 /product="Coproporphyrinogen III oxidase and related Fes
 oxidoreductases"
 /protein_id="CAD84189.1"
 /db_xref="GI:30138183"
 /translation="WPLVSPDPSKSPHIAELPPLGLYLHPWCVRKPCYCDNSHAIN
 SGGLDPESEVYVAKLELALPLVGRPVTVSPGCGTSLSATADAILSHVRML
 LPLAPQVEITLANFGTLEAKQFTDFHAGINRLSIGIQSNRHLQALGRIDGAE
 LRAEALQHFNSVLDLMYALPGFTDREALQDIETACNRGVHAHSYHUIETPFF
 YRYPQLPDESDAMQWMEIHLAERGYPHYETSAFAQPKPCLHMNTWQYGDYIG
 IGAGHASKISFRDRIIRQMRHKKHPQYLEQAGMAGNACLENLLATEQEVGPQDRIFPM
 MNALRIETGDFGLFHTRTGLTITIASIQKSLAEQRLIEWQHDICRPTPEGRFLND
 LLEIFLS"
 6979..8820
 /gene="pcoa"
 /locus_tag="NE0279"
 6979..8820
 /gene="pcoa"

Query Match 3.4%; Score 21; DB 1; Length 313050;

Best Local Similarity 100.0%; Pred. No. 2;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

427 GCCCGGGTACTGCTCAAAAT 447

Db 121608 GCCCGGGTACTGCTCAAAAT 121588

RESULT 5

MMU011080/c 2080 bp mRNA linear ROD 24-FEB-1999
 LOCUS Mus musculus mRNA for alpha-albumin protein.
 DEFINITION

ACCESSION AV011080

VERSION AV011080.1 GI:3646360

KEYWORDS alpha-albumin.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 Van Reeth, T., Gabant, P., Dreze, P., Szpirer, J. and Szpirer, C.

AUTHORS Unpublished

JOURNAL 2 (bases 1 to 2080)

REFERENCE Van Reeth, T.

AUTHORS Direct Submission

TITLE Submitted (15-SEP-1998)

JOURNAL Van Reeth T., departement de Biologie

moleculaire, Universite Libre de Bruxelles, laboratoire de Biologie

du Developpement, Rue des Chevaux, 57, 1640 Rhode St Genese,

BEELIUM

location/Qualifiers

1..2080

/organism="Mus musculus"

/mol_type="mRNA"

/db_xref="taxon:10090"

/chromosome="5"

/clone="IMAGE 522099"

/issue_type="diaphragm"

/clone_lib="Stratagene mouse diaphragm (#937303)"

/dev_stage="adult"

1..2080

/gene="alpha-albumin"

72..1907

/gene="alpha-albumin"

/codon_start=1

/product="alpha-albumin protein"

/protein_id="CAA09471.1"

/db_xref="GI:3646361"

/db_xref="GOA:O89020"

/translation="MRHLTGTGFIFFLLPLTESLALPTKPDVDHFNATQKIDENTA
 YLAIATSOYVQASDFVETLVKVLVDYDRCDWADNTLPESCKTANDAIQMLCDME
 GLPKHNFSCGKAGPRELCFYNNKANVGFPPPTLDPEKCOAYKNSGFLH
 LYNEVARRNPFPVPLVLAAMFEAAATCCQOQKATCFQAKAAPITQYLKASS
 YQNVGALIKFQKVLNSINAVFSKPKFQKLTLLLEDVSMYEGCGEGVYH
 CIRQSVQNHICSKQDSISSKIVCCCKTLREACTINANKDRPEGLSREKAKFT
 ESENVQERSDPKDFABFIYISRRHPLSTPRLRITKVMDFLEDCSRENPAG
 CYRVEDKFNETTQSLAVVQECNQFELGKDTLQHFVKTFAAPQPMEEVLVS
 SKENVAALITCCITLDEFACVADLVLGELGVNTRNTINPAVDHCKTDFAFRRH
 CFELKADTTVALPSVSLVSLATDWCOPKEDLQKXKRLVNLVKMMPGITDEEW
 LCLFTKTAARESCSEVQEPESCFSPESSKGTGDSQATENKITDQ"

ORIGIN

Query Match 3.4%; Score 20; DB 10; Length 2080;

Best Local Similarity 100.0%; Pred. No. 5.2;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 382 CCAGTTTGAACCTCTCTGG 401

Db 1864 CCAGTTTGAACCTCTCTGG 1845

RESULT 6

AC117443/c

LOCUS 22190 bp DNA linear PRI 26-SEP-2002

DEFINITION Homo sapiens 3 BAC RP11-208N8 (Roswell Park Cancer Institute Human

BAC Library) complete sequence.

ACCESSION AC117443

VERSION AC117443.4 GI:22024339
 KEYWORDS HTC.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

**REFERENCE
AUTHORS**

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 22190)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
Albrooks, S.L., Amaraturge, H.C., Are, J.R., Ayele, M., Banks, T.,
Barbaria, J., Benton, C., Blinige, K., Blankenburg, K., Bonnin, D.,
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant N.P.,
D.

Carron, T. F., Carter, M., Cavazos, S. R., Chacko, J., Chavez, D.,

Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleaveland, C.B., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, B., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dirh, H.H., Pouthwaite, K.J., Draber, H., Dugan-Rocha, S., Durbin, K.J.,

Earnhart, C., Edgar, D., Edwards, C. C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J. H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, N., Hawes, A., He, X., Hernandez, J., Hernandez, O., Hodgson, A., Hocues, M., Holloway, C.,

[illegible]

Matinez, R.; Massey, E.; Mawhinney, S.; Mubeco, M. P.; Meador, M.;
Niel, G.; Metzger, M.; Miner, G.; Miner, Z.; Mitchell, T.; Mohabbat, K.;
Moore, S.; Morgan, M.; Morrison, T.; Morris, S.; Moser, M.; Neal, D.;
Nelson, D.; Newton, J.; Newsham, N.; Nguyen, A.; Nguyen, N.; Nguyen, N.
Nickserson, E.; Nwokenkwo, S.; Ogih, M.; Okwuonu, G.; Otaguene, N.;
Oviedo, R.; Pace, A.; Pavton, B.; Peerv, J.; Perez, L.; Rangel, R.;

Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G., Scherer, S.

Scott, G., Shen, H., Shooshtari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A.

Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tasey, J., Taylor, C. C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalob, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczky, R., Wooten, S., Worley, K., Wu, C., Wu, Y.,

Wu, Y. F., Zhou, J., Zorrilla, S., Naylor, S. L., Weinstein, G. and Gibbs, R.

TITLE	Direct Submission	Unpublished
JOURNAL		

REFERENCE
2 (bases 1 to 22190)
AUTHORS
Worley, K. C.

JOURNAL TITLE Direct Submission
Submitted (10-APR-2002) Human Genome Sequencing Center. Department

Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 22190)
Worley, K. C.

TITLE	Direct Submission	
JOURNAL	Submitted (09-MAY-2002)	Human Genome Sequencing Center.

Submitted: 105 JAN 2004; Human Genome Sequencing Center, Baylor College of Medicine, One
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE
4 (bases 1 to 22190)
AUTHORS
Worley, K. C.

TITLE
Direct Submission
Submitted (31-JUL-2002) Human Genome Sequencing Center. Department

Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE
5 (bases 1 to 22190)
AUTHORS
Worley, K. C.

TITLE	Direct Submission	Department
Submitted (26-SEP-2002)	Human Genome Sequencing Center	

of Molecular and Human Genetics, Baylor College of Medicine, One

```

repeat_region      complement(8414..8473)
                    /rpt_family="HERVL"
repeat_region      8474..8796
                    /rpt_family="MR844A"
repeat_region      complement(8797..8913)
                    /rpt_family="HERVL"
repeat_region      complement(8919..9763)
                    /rpt_family="HERVL"
repeat_region      10824..11174
                    /rpt_family="THE1B"
repeat_region      complement(11511..11925)
                    /rpt_family="MLT1B"
repeat_region      12103..12131
                    /rpt_family="(CTA)n"
repeat_region      complement(13658..13691)
                    /rpt_family="U6"
repeat_region      14610..14747
                    /rpt_family="MLT1"
repeat_region      14827..15888
                    /rpt_family="L2"
repeat_region      15963..16008
                    /rpt_family="AT_rich"
repeat_region      16053..16120
                    /rpt_family="AT_rich"
repeat_region      16373..16475
                    /rpt_family="(TATATG)n"
repeat_region      complement(16816..17006)
                    /rpt_family="MR83C"
repeat_region      complement(17179..17300)
                    /rpt_family="MR83-internal"
repeat_region      17432..17489
                    /rpt_family="(TG)n"
repeat_region      complement(18943..19012)
                    /rpt_family="HUE8-P3"
repeat_region      complement(19013..19367)
                    /rpt_family="THE1B"
repeat_region      complement(19368..19835)
                    /rpt_family="HUE8-P3"
misc_feature       complement(20204..22190)
                    /note="overlaps bases 94715..96701 of clone AC128685"
                    /function="clone overlap"
repeat_region      complement(20701..21077)
                    /rpt_family="MR66-internal"
repeat_region      complement(21804..22190)
                    /rpt_family="MR83-internal"

ORIGIN
Query Match      3.3%; Score 20; DB 9; Length 22190;
Best Local Similarity 100.0%; Pred. No. 6.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      322 AAATGGCTGAATTCCTCA 341
        |||||
Db       21719 AAATGGCTGAATTCCTCA 21700

RESULT 7
AC138385
LOCUS      Mus musculus clone R224-383E18, LOW-PASS SEQUENCE SAMPLING.
DEFINITION
AC138385
VERSION    1 GI:27375071
KEYWORDS   HTG; HTGS PHASE0.
SOURCE     Mus musculus (house mouse)
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 86385)
AUTHORS     Birren,B., Nuebaum,C. and Lander,B.
TITLE       Mus musculus, clone RP24-383E18
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 86385)
AUTHORS     Birren,B., Nuebaum,C., Lander,B., Ali,A., Allen,N., Anderson,S.,

```

```

Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagos,B., Horton,A., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kellis,C., Landers,I., Levine,K.,
Lindblad-Toh,K., Liu,G., Maclean,C., MacDonald,P., Major,J.,
Matthews,C., McCarthy,M., Meidrim,J., Menes,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schuback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zaincun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (27-DEC-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L26666
Center clone name: 383_E_18
-----
* NOTE: This record contains 80 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 972: contig of 972 bp in length
* 973 1072: gap of 100 bp
* 1073 2033: contig of 961 bp in length
* 2034 2133: gap of 100 bp
* 2134 3004: contig of 871 bp in length
* 3005 3104: gap of 100 bp
* 3105 4089: contig of 985 bp in length
* 4090 4189: gap of 100 bp
* 4190 5167: contig of 978 bp in length
* 5168 5267: gap of 100 bp
* 5268 6264: contig of 997 bp in length
* 6265 6364: gap of 100 bp
* 6365 7349: contig of 985 bp in length
* 7350 7449: gap of 100 bp
* 7450 8407: contig of 958 bp in length
* 8408 8507: gap of 100 bp
* 8508 9478: contig of 971 bp in length
* 9479 9578: gap of 100 bp
* 9579 10554: contig of 976 bp in length
* 10555 10654: gap of 100 bp
* 10655 11636: contig of 982 bp in length
* 11637 11736: gap of 100 bp
* 11737 12693: contig of 957 bp in length
* 12694 12793: gap of 100 bp
* 12794 13293: contig of 936 bp in length
* 13293 13829: gap of 100 bp
* 13830 14828: contig of 999 bp in length
* 14829 14928: gap of 100 bp
* 14929 15916: contig of 988 bp in length
* 15917 16016: gap of 100 bp

```

16017 17012: contig of 996 bp in length
17013 17112: gap of 100 bp
17113 18087: contig of 975 bp in length
18088 18187: gap of 100 bp
18188 19181: contig of 994 bp in length
19182 19281: gap of 100 bp
19282 20257: contig of 976 bp in length
20258 20357: gap of 100 bp
20358 21402: contig of 1045 bp in length
21403 21502: gap of 100 bp
21503 22492: contig of 990 bp in length
22493 22592: gap of 100 bp
22593 23578: contig of 986 bp in length
23579 23678: gap of 100 bp
23679 24649: contig of 971 bp in length
24650 24749: gap of 100 bp
24750 25707: contig of 958 bp in length
25708 25807: gap of 100 bp
25809 26799: contig of 992 bp in length
26800 26899: gap of 100 bp
26900 27894: contig of 995 bp in length
27895 27994: gap of 100 bp
27996 29003: contig of 1009 bp in length
29004 29103: gap of 100 bp
29104 30120: contig of 1017 bp in length
30121 30220: gap of 100 bp
30221 31193: contig of 973 bp in length
31194 31293: gap of 100 bp
31294 32668: contig of 975 bp in length
32669 33353: contig of 985 bp in length
33354 33453: gap of 100 bp
33454 34379: contig of 926 bp in length
34380 34479: gap of 100 bp
34480 35464: contig of 985 bp in length
35465 35564: gap of 100 bp
35565 36537: contig of 973 bp in length
36538 36637: gap of 100 bp
36639 37628: contig of 991 bp in length
37629 37728: gap of 100 bp
37729 38721: contig of 993 bp in length
38722 38821: gap of 100 bp
38822 39817: contig of 996 bp in length
39818 39917: gap of 100 bp
39919 40919: contig of 1002 bp in length
40920 41019: gap of 100 bp
41020 41969: contig of 950 bp in length
41970 42069: gap of 100 bp
42070 43057: contig of 988 bp in length
43058 43157: gap of 100 bp
43158 44163: contig of 1006 bp in length
44164 44263: gap of 100 bp
44264 45276: contig of 1013 bp in length
45277 45376: gap of 100 bp
45377 46363: contig of 987 bp in length
46364 46463: gap of 100 bp
46464 47423: contig of 960 bp in length
47424 47523: gap of 100 bp
47524 48459: contig of 936 bp in length
48460 48559: gap of 100 bp
48560 49553: contig of 994 bp in length
49554 49653: gap of 100 bp
49654 50640: contig of 987 bp in length
50641 50740: gap of 100 bp
50741 51735: contig of 995 bp in length
51736 51835: gap of 100 bp
51836 52840: contig of 1005 bp in length
52841 52940: gap of 100 bp
52941 53926: contig of 996 bp in length
53927 54026: gap of 100 bp
54027 55022: contig of 996 bp in length
55023 55122: gap of 100 bp
55123 56111: contig of 989 bp in length

* 56112 56211: gap of 100 bp
56212 57188: contig of 977 bp in length
57189 57288: gap of 100 bp
57289 58253: contig of 965 bp in length
58254 58353: gap of 100 bp
58354 59263: contig of 910 bp in length
59264 59363: gap of 100 bp
59364 60378: contig of 1015 bp in length
60379 60478: gap of 100 bp
60479 61471: contig of 993 bp in length
61472 61571: gap of 100 bp
61572 62570: contig of 999 bp in length
62571 62670: gap of 100 bp
62671 63657: contig of 987 bp in length
63658 63757: gap of 100 bp
63759 64741: contig of 984 bp in length
64742 64841: gap of 100 bp
64842 65799: contig of 958 bp in length
65800 65899: gap of 100 bp
65900 66904: contig of 1005 bp in length
66905 67004: gap of 100 bp
67005 67983: contig of 979 bp in length
67984 68083: gap of 100 bp
68084 69072: contig of 989 bp in length
69073 69172: gap of 100 bp
69173 70169: contig of 997 bp in length
70170 70269: gap of 100 bp
70270 71267: contig of 998 bp in length
71268 71367: gap of 100 bp
71368 72375: contig of 1008 bp in length
72376 72475: gap of 100 bp
72476 73471: contig of 996 bp in length
73472 73571: gap of 100 bp
73572 74574: contig of 1003 bp in length
74575 74674: gap of 100 bp
74675 75670: contig of 996 bp in length
75671 75770: gap of 100 bp
75771 76750: contig of 980 bp in length
76751 76850: gap of 100 bp

Query Match 3.3%; Score 20; DB 2; Length 86385;

Best Local Similarity 100.0%; Pred. No. 7.4;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 124 ACTTGTCTCACTGCTCTTG 143

Db 8746 ACTTGTCTCACTGCTCTTG 8765

RESULT 8

AC138385/c

LOCUS

DEFINITION

AC138385

AC138385

VERSION

HTG; HTGS PHASE0.

KEYWORDS

SOURCE

ORGANISM

Mus musculus

REFERENCE

1 (bases 1 to 86385)

2 (bases 1 to 86385)

3 (bases 1 to 86385)

4 (bases 1 to 86385)

5 (bases 1 to 86385)

6 (bases 1 to 86385)

7 (bases 1 to 86385)

8 (bases 1 to 86385)

9 (bases 1 to 86385)

10 (bases 1 to 86385)

11 (bases 1 to 86385)

12 (bases 1 to 86385)

13 (bases 1 to 86385)

14 (bases 1 to 86385)

15 (bases 1 to 86385)

16 (bases 1 to 86385)

17 (bases 1 to 86385)

AC138385 Mus musculus clone RP24-383E18, LOW-PASS SEQUENCE SAMPLING. linear HTG 27-DEC-2002

AC138385 Mus musculus clone RP24-383E18, LOW-PASS SEQUENCE SAMPLING. linear HTG 27-DEC-2002

AC138385.1 GI:27375071

HTG; HTGS PHASE0.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 86385)

2 (bases 1 to 86385)

3 (bases 1 to 86385)

4 (bases 1 to 86385)

5 (bases 1 to 86385)

6 (bases 1 to 86385)

7 (bases 1 to 86385)

8 (bases 1 to 86385)

9 (bases 1 to 86385)

10 (bases 1 to 86385)

11 (bases 1 to 86385)

12 (bases 1 to 86385)

13 (bases 1 to 86385)

14 (bases 1 to 86385)

15 (bases 1 to 86385)

16 (bases 1 to 86385)

17 (bases 1 to 86385)

18 (bases 1 to 86385)

19 (bases 1 to 86385)

20 (bases 1 to 86385)

21 (bases 1 to 86385)

22 (bases 1 to 86385)

23 (bases 1 to 86385)

24 (bases 1 to 86385)

25 (bases 1 to 86385)

26 (bases 1 to 86385)

27 (bases 1 to 86385)

28 (bases 1 to 86385)

29 (bases 1 to 86385)

30 (bases 1 to 86385)

31 (bases 1 to 86385)

32 (bases 1 to 86385)

33 (bases 1 to 86385)

34 (bases 1 to 86385)

35 (bases 1 to 86385)

36 (bases 1 to 86385)

37 (bases 1 to 86385)

38 (bases 1 to 86385)

39 (bases 1 to 86385)

40 (bases 1 to 86385)

41 (bases 1 to 86385)

42 (bases 1 to 86385)

43 (bases 1 to 86385)

44 (bases 1 to 86385)

45 (bases 1 to 86385)

46 (bases 1 to 86385)

47 (bases 1 to 86385)

48 (bases 1 to 86385)

49 (bases 1 to 86385)

50 (bases 1 to 86385)

51 (bases 1 to 86385)

52 (bases 1 to 86385)

53 (bases 1 to 86385)

54 (bases 1 to 86385)

55 (bases 1 to 86385)

56 (bases 1 to 86385)

57 (bases 1 to 86385)

58 (bases 1 to 86385)

59 (bases 1 to 86385)

60 (bases 1 to 86385)

61 (bases 1 to 86385)

62 (bases 1 to 86385)

63 (bases 1 to 86385)

64 (bases 1 to 86385)

65 (bases 1 to 86385)

66 (bases 1 to 86385)

67 (bases 1 to 86385)

68 (bases 1 to 86385)

69 (bases 1 to 86385)

70 (bases 1 to 86385)

71 (bases 1 to 86385)

72 (bases 1 to 86385)

73 (bases 1 to 86385)

74 (bases 1 to 86385)

75 (bases 1 to 86385)

76 (bases 1 to 86385)

77 (bases 1 to 86385)

78 (bases 1 to 86385)

79 (bases 1 to 86385)

80 (bases 1 to 86385)

81 (bases 1 to 86385)

82 (bases 1 to 86385)

83 (bases 1 to 86385)

84 (bases 1 to 86385)

85 (bases 1 to 86385)

86 (bases 1 to 86385)

87 (bases 1 to 86385)

88 (bases 1 to 86385)

89 (bases 1 to 86385)

90 (bases 1 to 86385)

91 (bases 1 to 86385)

92 (bases 1 to 86385)

93 (bases 1 to 86385)

94 (bases 1 to 86385)

95 (bases 1 to 86385)

96 (bases 1 to 86385)

97 (bases 1 to 86385)

98 (bases 1 to 86385)

99 (bases 1 to 86385)

100 (bases 1 to 86385)

101 (bases 1 to 86385)

102 (bases 1 to 86385)

103 (bases 1 to 86385)

104 (bases 1 to 86385)

105 (bases 1 to 86385)

106 (bases 1 to 86385)

107 (bases 1 to 86385)

108 (bases 1 to 86385)

109 (bases 1 to 86385)

110 (bases 1 to 86385)

111 (bases 1 to 86385)

112 (bases 1 to 86385)

113 (bases 1 to 86385)

114 (bases 1 to 86385)

115 (bases 1 to 86385)

116 (bases 1 to 86385)

117 (bases 1 to 86385)

118 (bases 1 to 86385)

119 (bases 1 to 86385)

120 (bases 1 to 86385)

121 (bases 1 to 86385)

122 (bases 1 to 86385)

123 (bases 1 to 86385)

124 (bases 1 to 86385)

125 (bases 1 to 86385)

126 (bases 1 to 86385)

127 (bases 1 to 86385)

128 (bases 1 to 86385)

129 (bases 1 to 86385)

130 (bases 1 to 86385)

131 (bases 1 to 86385)

132 (bases 1 to 86385)

133 (bases 1 to 86385)

134 (bases 1 to 86385)

135 (bases 1 to 86385)

136 (bases 1 to 86385)

137 (bases 1 to 86385)

138 (bases 1 to 86385)

139 (bases 1 to 86385)

140 (bases 1 to 86385)

141 (bases 1 to 86385)

142 (bases 1 to 86385)

143 (bases 1 to 86385)

144 (bases 1 to 86385)

145 (bases 1 to 86385)

146 (bases 1 to 86385)

147 (bases 1 to 86385)

148 (bases 1 to 86385)

149 (bases 1 to 86385)

150 (bases 1 to 86385)

151 (bases 1 to 86385)

152 (bases 1 to 86385)

153 (bases 1 to 86385)

154 (bases 1 to 86385)

155 (bases 1 to 86385)

156 (bases 1 to 86385)

157 (bases 1 to 86385)

158 (bases 1 to 86385)

159 (bases 1 to 86385)

160 (bases 1 to 86385)

161 (bases 1 to 86385)

162 (bases 1 to 86385)

163 (bases 1 to 86385)

164 (bases 1 to 86385)

165 (bases 1 to 86385)

166 (bases 1 to 86385)

167 (bases

Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J.,
 Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T.,
 Mlenka, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
 Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
 Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R.,
 Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupbach, R.,
 Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K.,
 Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X.,
 Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (27-DEC-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE
 JOURNAL
 COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L26666
 Center clone name: 383_E_18

* NOTE: This record contains 80 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying relationships among clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

* 1
 * 972: contig of 972 bp in length
 * 973: gap of 100 bp
 * 1073: contig of 961 bp in length
 * 2034: gap of 100 bp
 * 2134: contig of 871 bp in length
 * 3005: gap of 100 bp
 * 3105: contig of 985 bp in length
 * 4090: gap of 100 bp
 * 4190: contig of 978 bp in length
 * 5168: gap of 100 bp
 * 5268: contig of 997 bp in length
 * 6265: gap of 100 bp
 * 6365: contig of 985 bp in length
 * 7350: gap of 100 bp
 * 7450: contig of 958 bp in length
 * 8408: gap of 100 bp
 * 8508: contig of 971 bp in length
 * 9479: gap of 100 bp
 * 9579: contig of 976 bp in length
 * 10555: gap of 100 bp
 * 10655: contig of 982 bp in length
 * 11637: gap of 100 bp
 * 11737: contig of 957 bp in length
 * 12694: gap of 100 bp
 * 12794: contig of 936 bp in length
 * 13730: gap of 100 bp
 * 13830: contig of 999 bp in length
 * 14829: gap of 100 bp
 * 14929: contig of 988 bp in length
 * 15917: gap of 100 bp
 * 16017: contig of 996 bp in length
 * 17013: gap of 100 bp
 * 17113: contig of 975 bp in length
 * 18088: gap of 100 bp
 * 18188: contig of 994 bp in length
 * 19181: gap of 100 bp
 * 19281: contig of 976 bp in length
 * 19282: contig of 976 bp in length
 * 20258: gap of 100 bp
 * 20358: contig of 1045 bp in length
 * 21402: gap of 100 bp
 * 21502: contig of 990 bp in length
 * 22492: gap of 100 bp
 * 22592: contig of 988 bp in length
 * 23578: gap of 100 bp
 * 23679: contig of 971 bp in length
 * 24649: gap of 100 bp
 * 24749: contig of 958 bp in length
 * 24750: gap of 100 bp
 * 25708: contig of 992 bp in length
 * 25807: gap of 100 bp
 * 25808: contig of 992 bp in length
 * 26800: gap of 100 bp
 * 26900: contig of 995 bp in length
 * 27895: gap of 100 bp
 * 27995: contig of 1009 bp in length
 * 29004: gap of 100 bp
 * 29104: contig of 1017 bp in length
 * 30121: gap of 100 bp
 * 30220: contig of 973 bp in length
 * 31193: gap of 100 bp
 * 31294: contig of 975 bp in length
 * 32269: gap of 100 bp
 * 32369: contig of 985 bp in length
 * 33354: gap of 100 bp
 * 33454: contig of 926 bp in length
 * 34380: gap of 100 bp
 * 34479: contig of 985 bp in length
 * 35464: gap of 100 bp
 * 35564: contig of 973 bp in length
 * 35638: gap of 100 bp
 * 36338: contig of 991 bp in length
 * 37629: gap of 100 bp
 * 37729: contig of 993 bp in length
 * 38722: gap of 100 bp
 * 38822: contig of 996 bp in length
 * 39818: gap of 100 bp
 * 39918: contig of 1002 bp in length
 * 40920: gap of 100 bp
 * 41020: contig of 950 bp in length
 * 41970: gap of 100 bp
 * 42070: contig of 988 bp in length
 * 43058: gap of 100 bp
 * 43158: contig of 1006 bp in length
 * 44164: gap of 100 bp
 * 44264: contig of 1013 bp in length
 * 45277: gap of 100 bp
 * 45377: contig of 987 bp in length
 * 46364: gap of 100 bp
 * 46464: contig of 960 bp in length
 * 47424: gap of 100 bp
 * 47524: contig of 936 bp in length
 * 48460: gap of 100 bp
 * 48560: contig of 994 bp in length
 * 49554: gap of 100 bp
 * 49654: contig of 987 bp in length
 * 50641: gap of 100 bp
 * 50741: contig of 995 bp in length
 * 51736: gap of 100 bp
 * 51836: contig of 1005 bp in length
 * 52841: gap of 100 bp
 * 52941: contig of 986 bp in length
 * 53927: gap of 100 bp
 * 54027: contig of 996 bp in length
 * 55023: gap of 100 bp
 * 55123: contig of 989 bp in length
 * 56112: gap of 100 bp
 * 56212: contig of 977 bp in length
 * 57189: gap of 100 bp
 * 57289: contig of 965 bp in length
 * 58254: gap of 100 bp
 * 58354: contig of 910 bp in length
 * 59264: gap of 100 bp
 * 59363: contig of 910 bp in length

```

* 59364 60378: contig of 1015 bp in length
* 60379 60478: gap of 100 bp
* 60479 61471: contig of 993 bp in length
* 61472 61571: gap of 100 bp
* 61572 62570: contig of 999 bp in length
* 62571 62670: gap of 100 bp
* 62671 63657: contig of 987 bp in length
* 63658 63757: gap of 100 bp
* 63758 64741: contig of 984 bp in length
* 64742 64841: gap of 100 bp
* 64842 65799: contig of 958 bp in length
* 65800 65899: gap of 100 bp
* 65900 66904: contig of 1005 bp in length
* 66905 67004: gap of 100 bp
* 67005 67983: contig of 979 bp in length
* 67984 68083: gap of 100 bp
* 68084 69072: contig of 989 bp in length
* 69073 69172: gap of 100 bp
* 69173 70169: contig of 997 bp in length
* 70170 70269: gap of 100 bp
* 70270 71267: contig of 998 bp in length
* 71268 71367: gap of 100 bp
* 71368 72375: contig of 1008 bp in length
* 72376 72475: gap of 100 bp
* 72476 73471: contig of 996 bp in length
* 73472 73571: gap of 100 bp
* 73572 74574: contig of 1003 bp in length
* 74575 74674: gap of 100 bp
* 74675 75670: contig of 996 bp in length
* 75671 75770: gap of 100 bp
* 75771 76750: contig of 980 bp in length
* 76751 76850: gap of 100 bp

Query Match
Best Local Similarity 3.3%; Score 20; DB 2; Length 86385;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 124 ACATTGCTCAACTGCTTGG 143
Db 80518 ACATTGCTCAACTGCTTGG 80499

RESULT 9
AC128685
LOCUS AC128685
DEFINITION Homo sapiens 3 BAC RP11-535124 (Roswell Park Cancer Institute Human
BAC library) complete sequence.
ACCESSION AC128685
VERSION AC128685.11 GI:32140198
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 121201)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
Alsbrooks, S.L., Amarantunge, H.C., Are, J.R., Ayale, M., Banks, T.,
Barbaria, J., Benton, J., Bimge, K., Blankenburg, K., Bonnin, D.,
Bouck, J., Bowie, S., Brieva, W., Brown, E., Brown, M., Bryant, N.P.,
Buhay, C., Burck, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carroll, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Douthwaite, K.J., Draper, K., Dugan-Rocha, S., Durbin, K.J.,
Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., He, X.,
Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C.,
Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,

```

```

Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,
Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
Loulseghe, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M.,
Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
Moore, S., Morgan, M., Moorish, T., Morris, S., Moser, M., Neal, D.,
Nelson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N.,
Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G., Oragunye, N.,
Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L.,
Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, I., Rivers, M.,
Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S.,
Scott, G., Shen, H., Shooshitari, N., Sisson, I., Sodergren, E.,
Sonaika, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A.,
Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C.,
Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L.,
Vera, V., Villalon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S.,
Warren, R., Washington, C., Watlington, S., Williams, G.,
Williamson, A., Wleciyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y.,
Wu, Y.F., Zhou, J., Zorrilla, S., Naylor, S.L., Weinstein, G. and
Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 121201)
Worley, K.C.
Direct Submission
Submitted (22-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 121201)
Worley, K.C.
Direct Submission
Submitted (13-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 121201)
Worley, K.C.
Direct Submission
Submitted (25-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 121201)
Worley, K.C.
Direct Submission
Submitted (26-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
6 (bases 1 to 121201)
Worley, K.C.
Direct Submission
Submitted (21-JUN-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jun 21, 2003 this sequence version replaced gi:23307944.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST

```

(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases.

Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:

<http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

FEATURES

source

Location/Qualifiers
 1. .121201
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="3"
 /clone="RP11-535124"
 1. .2094
 /note="overlaps bases 176428..178521 of clone AC112644"
 /function="clone overlap"

misc_feature

repeat_region

1. .499
 /rpt_family="LIMA3"

repeat_region

510..723
 /rpt_family="LIMA3"

repeat_region

735..2212
 /rpt_family="LIMA3"

repeat_region

2212..3980
 /rpt_family="LIMA3"

repeat_region

3981..4016
 /rpt_family="(TTTA)n"

repeat_region

complement(4017..4168)
 /rpt_family="LIPAB"

repeat_region

4169..4349
 /rpt_family="LIMA3"

repeat_region

4691..4720
 /rpt_family="AT_rich"

repeat_region

4756..4789
 /rpt_family="AT_rich"

repeat_region

5854..5902
 /rpt_family="AT_rich"

repeat_region

6406..6462
 /rpt_family="AT_rich"

repeat_region

complement(7193..8279)
 /rpt_family="MER52A"

repeat_region

complement(8280..8524)
 /rpt_family="LIPB1"

repeat_region

8525..8671
 /rpt_family="(TA)n"

repeat_region

complement(8672..10086)
 /rpt_family="LIPB1"

repeat_region

10087..10516
 /rpt_family="LIPB1"

repeat_region

complement(10517..11662)
 /rpt_family="LIPB1"

repeat_region

11659..11801
 /rpt_family="LIMA4"

repeat_region

12308..12628
 /rpt_family="THE1B"

repeat_region

12629..12693
 /rpt_family="THE1B-int"

repeat_region

12802..13984
 /rpt_family="THE1B-int"

repeat_region

14730..15081

repeat_region
 15082..15435
 /rpt_family="THE1B"
 repeat_region
 15971..16501
 /rpt_family="LIM4"
 repeat_region
 16514..16799
 /rpt_family="LIM4"
 repeat_region
 complement(16812..17269)
 /rpt_family="LIPAB"
 repeat_region
 17270..17614
 /rpt_family="LIM4"
 repeat_region
 18192..18274
 /rpt_family="MER41B"
 repeat_region
 18720..18761
 /rpt_family="(TG)n"
 repeat_region
 complement(18949..19480)
 /rpt_family="LIMD3"
 repeat_region
 20297..20335
 /rpt_family="AT_rich"
 repeat_region
 20506..20802
 /rpt_family="Tigger1"
 repeat_region
 21358..21407
 /rpt_family="AT_rich"
 repeat_region
 23313..23495
 /rpt_family="MIRB"

Query Match 3.3%; Score 20; DB 9; Length 121201;

Best Local Similarity 100.0%; Pred. NO. 7.6;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 322 AATGGCTGAATTCCTCA 341

Db 95186 AATGGCTGAATTCCTCA 95205

RESULT 10

AC090177/c

LOCUS

DEFINITION Homo sapiens chromosome 15 clone RP11-762M18 map 15, WORKING DRAFT

SEQUENCE, 8 unordered pieces.

AC090177.3 GI:22122959

VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 154934)

Birren, B., Nussbaum, C. and Lander, E.

Homo sapiens chromosome 15, clone RP11-762M18

2 (bases 1 to 154934)

Birren, B., Linton, L., Nussbaum, C., Lander, E., Allen, N., Anderson, S.,

Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B., Brown, A.,

Canarata, J., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S.,

Collamore, A., Cooke, P., DeAtellano, K., Dewar, K., Diaz, J.S.,

Dodge, S., Fero, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J.,

Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N.,

Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,

Jones, C., Karatas, A., Lacombe, K., Lamazares, R., Landers, T.,

Lehoczky, J., Levine, R., Liu, G., Maclean, C., Macdonald, P.,

Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K.,

McPheeters, R., Melidrim, J., Meneus, L., Mihova, T., Mienna, V.,

Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H.,

O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,

Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Ratta, R.,

Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M.,

Sougnaz, C., Spencer, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P.,

Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,

Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A.,

Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zaincun, J.,

Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
REFERENCE
AUTHORS

Direct Submission
Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 154934)
Birnbaum, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J. S., Dodge, S., Farc, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kellis, C., Lander, S., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Teeffaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 6, 2002 this sequence version replaced gi:13958454.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L12158
Center clone name: 762_M18
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 152208 bases at least Q40
Consensus quality: 153474 bases at least Q30
Consensus quality: 153909 bases at least Q20
Insert size: 157000; agarose-fp
Insert size: 154234; sum-of-contigs
Quality coverage: 8.6 in Q20 bases; agarose-fp
Quality coverage: 8.8 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

1 4950: contig of 4950 bp in length
* 4951 5050: gap of 100 bp
* 5051 5258: contig of 208 bp in length
* 5259 5359: gap of 100 bp
* 5359 12316: contig of 6958 bp in length
* 12317 12416: gap of 100 bp
* 12417 28647: contig of 16231 bp in length
* 28648 42560: contig of 13813 bp in length
* 42561 66512: gap of 100 bp
* 66513 101249: contig of 23852 bp in length
* 101250 101349: gap of 100 bp
* 101350 154934: contig of 34637 bp in length.
Location/Qualifiers

TITLE
JOURNAL
COMMENT

Source

1. .154934
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone_lib="RPC1-11 Human Male BAC"
1. 4950
/note="assembly_fragment"
clone_end:SP6
vector_side:left
5051. 5258
/note="assembly_fragment"
5359. 12316
/note="assembly_fragment"
12417. 28647
/note="assembly_fragment"
28748. 42560
/note="assembly_fragment"
42661. 66512
/note="assembly_fragment"
66613. 101249
/note="assembly_fragment"
101350. 154934
/note="assembly_fragment"
clone_end:T7
vector_side:right

ORIGIN

Query Match 3.3%; Score 20; DB 2; Length 154934;
Best Local Similarity 100.0%; Fred. No. 7.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 497 ACAGTGAGCTAGTCATAGAA 516
|||||
DB 28480 ACAGTGAGCTAGTCATAGAA 28461
|||||

RESULT 11
AC090181/c
LOCUS
AC090181
DEFINITION
AC090181
VERSION
AC090181.9 GI:27777610
KEYWORDS
HTG.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 157793)
AUTHORS
Birnbaum, B., Nusbaum, C. and Lander, E.
TITLE
Homo sapiens chromosome 15, clone RP11-797A18
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 157793)
AUTHORS
Birnbaum, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Choepel, Y., Colangeo, M., Collins, S., Collymore, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J. S., Dodge, S., Farc, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Karatas, A., LaRocque, K., Lamazares, R., Lander, S., Lehoczy, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Sougnuez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Teeffaye, S., Theodore, J.,

Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 171061)

Barren, B., Nusbaum, C., Lander, E., Abouelheil, A., Allen, N., Anderson, M., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguski, L., Boukhalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArillano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, X., Liu, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meidrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Risse, C., Rogov, P., Romar, J., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (15-DEC-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Dec 15, 2003 this sequence version replaced gl:38524692.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center -----

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information -----

Center project name: L20923

Center clone name: 176_C_2

* NOTE: This is a 'working draft' sequence. It currently

* consists of 2 contigs. Gaps between the contigs

* are represented as runs of N. The order of the pieces

* is believed to be correct as given, however the sizes

* of the gaps between them are based on estimates that have

* provided by the submitter.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

* 1 2589: contig of 2589 bp in length

* 2590 171061: gap of 100 bp

* 2690 171061: contig of 168372 bp in length.

Location/Qualifiers

1..171061

/organism="Mus musculus"

/mol_type="genomic DNA"

/db_xref="taxon:10090"

/chromosome="16"

/map="16"

/clone="RP23-176C2"

/clone_lib="RPCI-23 Female Mouse BAC"

Query Match 3.3%; Score 20; DB 2; Length 171061;

Best Local Similarity 100.0%; Pred. No. 7.9;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 124 ACTTTGCTCAACTGTCCTGG 143

|||||

Db 144646 ACTTTGCTCAACTGTCCTGG 144665

AC005900

LOCUS

RESULT 13

AC005900/c

linear PRI 06-NOV-1999

Mon Aug 2 09:36:50 2004

```

DEFINITION Homo sapiens chromosome 17, clone hRPK.998_F_8, complete sequence.
ACCESSION AC005900
VERSION AC005900.1 GI:3849824
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 175066)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 17, clone hRPK.998_F_8
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 175066)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C.,
Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S.,
Collins,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R.,
Gage,D., Gardyna,S., Geraghty,K., Grant,G., Hagos,B., Heaford,A.,
Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L.,
Karatas,A., Lehoczy,J., Macdonald,P., Marquis,N., McEwan,P.,
McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J.,
Mychaleckyj,J., Nafé,R., Naylor,J., Niloff,M., O'Connor,T.,
O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,
Severy,P., Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I.,
Vasiliiev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
Ye,W.J., Zhao,J. and Zody,M.
DIRECT SUBMISSION
Submitted (31-Oct-1998) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 175066)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C.,
Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S.,
Collins,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R.,
Gage,D., Gardyna,S., Geraghty,K., Grant,G., Hagos,B., Heaford,A.,
Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L.,
Karatas,A., Lehoczy,J., Macdonald,P., Marquis,N., McEwan,P.,
McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J.,
Mychaleckyj,J., Nafé,R., Naylor,J., Niloff,M., O'Connor,T.,
O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,
Severy,P., Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I.,
Vasiliiev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
Ye,W.J., Zhao,J. and Zody,M.
DIRECT SUBMISSION
Submitted (06-NOV-1998) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Nov 6, 1998 this sequence version replaced gi:3845400.
All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES
source
1..175066
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="hRPK.998_F_8"
/clone_lib="RPCL-17 human BAC library"
899..1372
/rpt_family="L2"
repeat_region
complement(1547..1577)
/rpt_family="CA)n"
repeat_region
1922..2133
/rpt_family="L2"
repeat_region
complement(2217..2249)
/rpt_family="CA)n"
repeat_region
4282..4579
/rpt_family="AluSc"
repeat_region
complement(5098..5129)
/rpt_family="CAAA)n"
repeat_region
complement(5130..5433)
/rpt_family="AluY"
repeat_region
complement(5922..6100)
/rpt_family="L1PA13"
6101..6182
/rpt_family="L1PI"
6327..6435
/rpt_family="L1PI"
7005..7302
/rpt_family="AluX"
7303..7613
/rpt_family="AluSp"
7615..7691
/rpt_family="A-rich"
complement(8284..9314)
/rpt_family="MER4-internal"
10996..11702
/rpt_family="LTR8"
complement(12089..12397)
/rpt_family="AluSp"
complement(12536..12692)
/rpt_family="AluSp/g"
complement(12770..13056)
/rpt_family="MER57-internal"
complement(13069..13166)
/rpt_family="HERV23"
complement(13194..13290)
/rpt_family="MER57-internal"
complement(13487..13591)
/rpt_family="MER4-internal"
complement(13629..13698)
/rpt_family="L1M4"
complement(13699..13738)
/rpt_family="TAAA)n"
complement(13739..14018)
/rpt_family="AluSg"
complement(14019..14066)
/rpt_family="L1M4"
complement(14067..14088)
/rpt_family="CA)n"
complement(14089..14209)
/rpt_family="L1M4"
complement(14793..15178)
/rpt_family="L1M4"
complement(15281..19632)
/rpt_family="L1PA2"
complement(19644..19752)
/rpt_family="L1P"
complement(19760..20454)
/rpt_family="L1ME1"
complement(20455..20615)
/rpt_family="AluY"
20616..20846
/rpt_family="TA)n"
complement(20847..20784)
/rpt_family="AluY"
complement(20785..21187)
/rpt_family="L1ME1"
complement(21208..21386)
/rpt_family="MIR"
21387..21688
/rpt_family="AluY"
complement(21689..21743)
/rpt_family="MIR"
complement(21821..21907)
/rpt_family="L1P"
22235..22536
/rpt_family="LMC4"
23987..24152
/rpt_family="L2"
24233..24516

```

```

/rpt_family="Tigger2"
24517 .24634
/rpt_family="AluSq/x"
24663 .25126
/rpt_family="Tigger2"
complement(25555..25430)
/rpt_family="MER50"
complement(25449..25627)
/rpt_family="AluSq/x"
complement(25629..26179)
/rpt_family="MER50"
26205 26465
/rpt_family="Tigger2"
complement(26829..26951)
/rpt_family="purine-rich"
complement(26966..27083)
/rpt_family="purine-rich"
27097 .27174
/rpt_family="CCGTTn"
complement(27137..27256)
/rpt_family="G-rich"
complement(27227..27354)
/rpt_family="GGGAA)n"
complement(27329..27432)
/rpt_family="purine-rich"
complement(27850..28742)
/rpt_family="LiM4"
28882 .28932
/rpt_family="AT rich"
complement(29470..29778)
/rpt_family="AluSg1"
30002 .30029
/rpt_family="AT rich"
complement(31018..31069)
/rpt_family="LiMB3"
31070 .31140
/rpt_family="(TA)n"
complement(31148..31164)
/rpt_family="LiMB3"
complement(31165..31462)
/rpt_family="AluY"
complement(31463..31787)
/rpt_family="LiMB3"
complement(31788..32098)
/rpt_family="AluY"
complement(32099..32179)
/rpt_family="LiMB3"
complement(32274..32462)
/rpt_family="MER5A"
complement(33822..33952)
/rpt_family="LiPA5"
complement(35003..35313)
/rpt_family="AluYa5"

Query Match 3.3%; Score 20; DB 9; Length 175066;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 CTCAGTCATAAGTTTACC 77
Db 76063 CTCAGTCATAAGTTTACC 76044

RESULT 14
AC115834
LOCUS
DEFINITION Mus musculus clone RP24-15911, WORKING DRAFT SEQUENCE, 6 ordered
pieces
AC115834
VERSION AC115834.4 GI:32567880
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

```

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 180172)

Birren, B., Nusbaum, C. and Lander, E. Unpublished

2 (bases 1 to 180172)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Bouckhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Collangelo, M., Collins, S., Collamore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fero, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hago, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Lacroque, K., Lamazares, R., Landers, T., Lenoczky, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triglio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, N. J., Young, G., Zairoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 180172)

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, M., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Bouckhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collamore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J. S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hago, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mieng, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zairoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (14-JUL-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jul 14, 2003 this sequence version replaced gi:21362168.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L24412

Center clone name: 159 L 1

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 177938 bases at least Q40

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 30, 2004, 14:14:35 / Search time 454 Seconds
(without alignments)
5698.574 Million cell updates/sec

Title: US-09-868-338-7_COPY_1117_1725

Perfect score: 609

Sequence: 1 atgtagaataatcaatgacct.....cggataccattatcaaaacta 609

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_29Jan04.*

- 1: Geneseqn1980s.*
- 2: Geneseqn1990s.*
- 3: Geneseqn2000s.*
- 4: Geneseqn2001as.*
- 5: Geneseqn2001bs.*
- 6: Geneseqn2002s.*
- 7: Geneseqn2003as.*
- 8: Geneseqn2003bs.*
- 9: Geneseqn2003cs.*
- 10: Geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	609	100.0	2370	3	AAA65395
C 2	30	4.9	30	3	AAA65394 Brevibact
C 3	22	3.6	22	3	AAA65393 Brevibact
C 4	19	3.1	768	7	ACN21027
C 5	19	3.1	855	8	ADA29053
C 6	18	3.0	446	4	AA522742 Human cdn
C 7	18	3.0	471	4	AA522506
C 8	18	3.0	496	8	ACH25483
C 9	18	3.0	498	8	ACH28170 Human adu
C 10	18	3.0	621	6	ABN92747
C 11	18	3.0	1125	4	ABL26531 Drosophil
C 12	18	3.0	2397	7	ADA00895 Mouse osc
C 13	18	3.0	3125	4	ABL26530 Drosophil
C 14	18	3.0	11920	4	ABL21028 Drosophil
C 15	17	2.8	60	6	ABN35170 Human spl
C 16	17	2.8	192	6	ABN35170 Human spl
C 17	17	2.8	264	5	AAH65860 C. glutami
C 18	17	2.8	376	4	AAH34551 Human col
C 19	17	2.8	439	7	ABX43841 Bovine ES
C 20	17	2.8	495	6	ABK74808 Bacillus
C 21	17	2.8	687	5	AAH65859 C. glutami
C 22	17	2.8	687	7	ACA00221 C. Glutam
C 23	17	2.8	700	3	AAA12925 Human alp

ALIGNMENTS

RESULT 1

AAA65395

ID AAA65395 standard; DNA; 2370 BP.

AC AAA65395;

DT 15-SEP-2003 (revised)

DT 09-NOV-2000 (first entry)

DE Brevibacterium lactofermentum gltBD gene DNA sequence SEQ ID NO:7.

XX Brevibacterium lactofermentum; ABC transporter; breeding; ATPase;

XX Corynebacterium; gltBD; L-glutamic acid; GS.

OS Corynebacterium glutamicum.

XX Key Location/Qualifiers

FT CDS 1..1104

FT /*tag= a

FT /product= "ORF1"

FT 1117..1128

FT /*tag= b

FT /product= "ORF2"

FT 1759..2370

FT /*tag= c

FT /product= "ORF3"

WO2000037647-A1.

29-JUN-2000.

16-DEC-1999; 99WO-JP007079.

18-DEC-1998; 98JP-00360621.

(AJIN) AJINOMOTO CO INC.

Kanno S, Kimura E, Matsui K, Nakamatsu T;

WPI; 2000-452189/39.

P-PSDB; AAB12591, AAB12592, AAB12593.

ABC transporter constituent of Brevibacterium lactofermentum, its encoded gene and variants, applicable in breeding Corynebacteria particularly for production of L-glutamic acid.

XX

PS Claim 14; Page 23-28; 34pp; Japanese.

XX The present invention describes a protein (I) which can be used to
 CC construct an ATP binding cassette (ABC) transporter. ABC transporter DNAs
 CC can be used in breeding Corynebacteria particularly for production of L-
 CC glutamic acid. The present sequence encodes three ORFs (open reading
 CC frames) from the Brevibacterium lactofermentum gltBD gene, which is used
 CC in the exemplification of the present invention. (Updated on 15-SEP-2003
 CC to standardise OS field)

XX Sequence 2370 BP; 633 A; 563 C; 515 G; 659 T; 0 U; 0 Other;

Query Match 100.0%; Score 609; DB 3; Length 2370;
 Best Local Similarity 100.0%; Pred. No. 8.7e-310;
 Matches 609; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGATGAAATCAATGACCTCAAGAAATCTTTGGCGTTTCGGATCTTATGGCAAGGTCCTC 60
 Db 1117 ATGATGAAATCAATGACCTCAAGAAATCTTTGGCGTTTCGGATCTTATGGCAAGGTCCTC 1176

Qy 61 AGTCATAAGTTTACAGAAATGACGACACTGACGAGCGTCCGGTTTCAGGAAA 120
 Db 1177 AGTCATAAGTTTACAGAAATGACGACACTGACGAGCGTCCGGTTTCAGGAAA 1236

Qy 121 TCGACTTTGCTCAACTGCTTGGCAGCTTGAACCAAGTTTCGGACAGATCTTGTGTC 180
 Db 1237 TCGACTTTGCTCAACTGCTTGGCAGCTTGAACCAAGTTTCGGACAGATCTTGTGTC 1296

Qy 181 GAGGATGTAGACCTTCTGAAATCTCTACGCGTAAGCAACGGTTATACAGGAAAAATACG 240
 Db 1297 GAGGATGTAGACCTTCTGAAATCTCTACGCGTAAGCAACGGTTATACAGGAAAAATACG 1356

Qy 241 GTGGGCTATTATTCAAGATTATGCTTGATTCGGACAGACAGTTAAATCAACCTT 300
 Db 1357 GTGGGCTATTATTCAAGATTATGCTTGATTCGGACAGACAGTTAAATCAACCTT 1416

Qy 301 CAGCTTGGCGTGAAAAACACAAATGGCGTGAATTCCTCAAGTACTTTCATGCTGTTGGT 360
 Db 1417 CAGCTTGGCGTGAAAAACACAAATGGCGTGAATTCCTCAAGTACTTTCATGCTGTTGGT 1476

Qy 361 CTTGAGTCTGTCGAGAAAGCAGTTTGTGAACTCTCTGTTGGCGAACAACACGAACT 420
 Db 1477 CTTGAGTCTGTCGAGAAAGCAGTTTGTGAACTCTCTGTTGGCGAACAACACGAACT 1536

Qy 421 GCGTTGGCCCGGTACTGCTCAAAAATCCCGAATAATTCCTGGCTGATGAACCAACCGGA 480
 Db 1537 GCGTTGGCCCGGTACTGCTCAAAAATCCCGAATAATTCCTGGCTGATGAACCAACCGGA 1596

Qy 481 GCTCTAGATTTAAACAAAGTAGCTAGTATAGAGCAATTCAGAGCACTCGCCGACAAA 540
 Db 1597 GCTCTAGATTTAAACAAAGTAGCTAGTATAGAGCAATTCAGAGCACTCGCCGACAAA 1656

Qy 541 GGGCGCACCGTTGTTGCTTACGACACTCGCCCTCTTCCGAGATCAGCGGATACCAT 600
 Db 1657 GGGCGCACCGTTGTTGCTTACGACACTCGCCCTCTTCCGAGATCAGCGGATACCAT 1716

Qy 601 ATCAAACTA 609
 Db 1717 ATCAAACTA 1725

RESULT 2
 AAA65394/c
 ID AAA65394 standard; DNA; 30 BP.
 XX
 AC AAA65394;
 XX
 XX 15-SEP-2003 (revised)
 DT 09-NOV-2000 (first entry)
 XX
 DE Brevibacterium lactofermentum gltBD gene PCR primer SEQ ID NO:6.
 XX
 XX Brevibacterium lactofermentum; ABC transporter; breeding; ATPase;

KW corynebacterium; gltBD; L-glutamic acid; PCR primer; ss.
 XX
 OS Corynebacterium glutamicum.
 PN WO200037647-A1.
 XX
 XX 29-JUN-2000.
 PD
 XX 16-DEC-1999; 99WO-JP007079.
 PF
 XX 18-DEC-1998; 98JP-00360621.
 PR
 XX (AJIN) AJINOMOTO CO INC.
 PA
 XX Kanno S, Kimura E, Matsui K, Nakamatsu T;
 PI
 XX WPI; 2000-452189/39.
 DR
 XX ABC transporter constituent of Brevibacterium lactofermentum, its encoded
 PT gene and variants, applicable in breeding Corynebacteria particularly for
 PT production of L-glutamic acid.
 PS Disclosure; Page 22; 34pp; Japanese.
 XX
 CC The present invention describes a protein (I) which can be used to
 CC construct an ATP binding cassette (ABC) transporter. ABC transporter DNAs
 CC can be used in breeding Corynebacteria particularly for production of L-
 CC glutamic acid. The present sequence represents a PCR primer for the
 CC Brevibacterium lactofermentum gltBD gene, which is used in the
 CC exemplification of the present invention. (Updated on 15-SEP-2003 to
 CC standardise OS field)
 XX
 SQ Sequence 30 BP; 7 A; 5 C; 10 G; 8 T; 0 U; 0 Other;

Query Match 4.9%; Score 30; DB 3; Length 30;
 Best Local Similarity 100.0%; Pred. No. 3.5e-05;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 129 GCTCAACTGCTTGGCACACTTGACAAACC 158
 Db 30 GCTCAACTGCTTGGCACACTTGACAAACC 1

RESULT 3
 AAA65393/c
 ID AAA65393 standard; DNA; 22 BP.
 XX
 AC AAA65393;
 XX
 DT 15-SEP-2003 (revised)
 DT 09-NOV-2000 (first entry)
 XX
 DE Brevibacterium lactofermentum gltBD gene PCR primer SEQ ID NO:5.
 XX
 KW Brevibacterium lactofermentum; ABC transporter; breeding; ATPase;
 KW corynebacterium; gltBD; L-glutamic acid; PCR primer; ss.
 XX
 OS Corynebacterium glutamicum.
 PN WO200037647-A1.
 XX
 XX 29-JUN-2000.
 PD
 XX 16-DEC-1999; 99WO-JP007079.
 PF
 XX 18-DEC-1998; 98JP-00360621.
 PR
 XX (AJIN) AJINOMOTO CO INC.
 PA
 XX Kanno S, Kimura E, Matsui K, Nakamatsu T;
 PI
 XX WPI; 2000-452189/39.
 DR
 XX

PT ABC transporter constituent of Brevibacterium lactofermentum, its encoded
 PT gene and variants, applicable in breeding Corynebacteria particularly for
 PT production of L-glutamic acid.

XX Disclosure; Page 21; 34pp; Japanese.

CC The present invention describes a protein (I) which can be used to
 CC construct an ATP binding cassette (ABC) transporter. ABC transporter DNAs
 CC can be used in breeding Corynebacteria particularly for production of L-
 CC glutamic acid. The present sequence represents a PCR primer for the
 CC Brevibacterium lactofermentum gltBD gene, which is used in the
 CC exemplification of the present invention. (Updated on 15-SEP-2003 to
 CC standardise OS field)

XX Sequence 22 BP; 5 A; 7 C; 5 G; 5 T; 0 U; 0 Other;

Query Match 3.6%; Score 22; DB 3; Length 22;

Best Local Similarity 100.0%; Pred. No. 0.56; Length 22;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 CGGACAGATCTTCTCGAGGAT 186

Db 22 CGGACAGATCTTCTCGAGGAT 1

RESULT 4

ADA21027

ID ACA21027 standard; DNA; 768 BP.

AC ACA21027;

DT 19-JUN-2003 (first entry)

DE Prokaryotic essential gene #2684.

XX Antisense; ds; prokaryotic essential gene; cell proliferation;
 KW drug design; gene.

OS Acinetobacter baumannii.

XX W0200277183-A2.

PD 03-OCT-2002.

PF 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

PI WPI; 2003-029926/02.

DR P-PSDB; ABU17157.

XX New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.

XX Claim 14; SEQ ID NO 8897; 1766pp; English.

CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the

CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway;
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
 CC prokaryotic essential genes. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 768 BP; 223 A; 149 C; 170 G; 226 T; 0 U; 0 Other;

Query Match

3.1%; Score 19; DB 7; Length 768;

Best Local Similarity 100.0%; Pred. No. 20;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 115 GGAAAATCGACTTTGCTCA 133

Db 115 GGAAAATCGACTTTGCTCA 133

RESULT 5

ADA29053

ID ADA29053 standard; DNA; 855 BP.

XX ADA29053;

DT 20-NOV-2003 (first entry)

DE DNA encoding Acinetobacter baumannii protein #340.

XX ds; Gene; Acinetobacter baumannii; bacterial disease; antibacterial;
 KW vaccine; plant biocontrol agent.

OS Acinetobacter baumannii.

XX US6562958-B1.

XX 13-MAY-2003.

XX 04-JUN-1999; 99US-00328352.

XX 09-JUN-1998; 98US-0088701P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Breton G, Bush D;

XX WPI; 2003-576092/54.

XX P-PSDB; ADA33179.

XX New Acinetobacter baumannii proteins and nucleic acids, useful as reagents
 PT for diagnosing a bacterial disease, as components of antibacterial
 PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for
 PT plants.

PS Example; SEQ ID NO 340; 328pp; English.

CC The invention relates to isolated Acinetobacter baumannii nucleic acids.
 CC The A. baumannii nucleic acids and polypeptides are useful as reagents
 CC for diagnosing a bacterial disease, as components of antibacterial
 CC vaccines, as targets for antibacterial drugs, to detect the presence of
 CC A. baumannii and other Acinetobacter species in a sample, in screening
 CC compounds for the ability to interfere with the A. baumannii life cycle
 CC or to inhibit A. baumannii infection, and as biocontrol agents for
 CC plants. The present sequence represents DNA encoding an A. baumannii
 CC protein.
 CC
 XX
 SQ Sequence 855 BP; 239 A; 168 C; 193 G; 255 T; 0 U; 0 Other;
 Query Match 3.1%; Score 19; DB 8; Length 855;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 115 GGAAATCGACTTGTCTCA 133
 Db 199 GGAAATCGACTTGTCTCA 217
 RESULT 6
 AAS22742/C
 ID AAS22742 standard; cDNA; 446 BP.
 XX AC AAS22742;
 XX DT 24-OCT-2001 (first entry)
 XX DE Human cDNA encoding a novel human protein #308.
 XX KW Human; novel protein; ss; Antianaemic; osteopathic; antiinflammatory;
 XX KW immunomodulatory; cytostatic; neuroprotective; vulnery; nootropic;
 XX KW anticonvulsant; antiallergic; cerebroprotective; antifungal; antiviral;
 XX KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
 XX KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
 XX KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
 XX KW tissue regeneration; immune disorder.
 XX OS Homo sapiens.
 XX WO200155437-A2.
 XX PN 25-JAN-2001; 2001WO-US002623.
 XX PD 02-AUG-2001.
 XX PR 25-JAN-2000; 2000US-00491404.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Tang YT, Liu C, Drmanac RT;
 XX WPI: 2001-451939/48.
 XX P-PSDB; AAU14437.
 XX PT Isolated polypeptides useful for treating anti-inflammatory diseases,
 XX nervous system disorders, and for regenerating bone and cartilage.
 XX PS Claim 1; Page 706; 894pp; English.
 XX
 CC The invention relates to polynucleotides encoding novel human proteins or
 CC their active domains. The polypeptides, polynucleotides and antibodies
 CC raised against the polypeptides are used in a method of treatment of a
 CC mammal and prevention of disorders caused by the aberrant protein
 CC expression or activity. The polypeptides can be used as molecular weight
 CC markers, food supplements, and in antibody production. The polypeptides
 CC are used to identify compounds which bind to the polypeptides.
 CC Polynucleotides of the invention are used as probes and primers, for
 CC sequencing, for chromosome or gene mapping, in the production of
 CC recombinant proteins, and in generating anti-sense DNA or RNA and in gene
 CC therapy. Polypeptides of the invention can be used to target drugs to a
 CC tumour, in assays to determine biological activity, to raise

CC antibodies/ elicit an immune response, to determine quantitative protein
 CC levels, as tissue markers, and to isolate receptors or ligands.
 CC Polypeptides of the invention may also be useful in treating platelet
 CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
 CC ligament and/or nerve tissue, wound healing, treating burns, promoting
 CC the proliferation, differentiation and survival of stem cells, as a
 CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
 CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
 CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
 CC fungal infection or from autoimmunity, cancer, allergy, asthma, graft-
 CC versus-host disease, eczema, haemophilia, thrombosis, anti-inflammatory
 CC diseases, nervous system disorders, and infection. The present sequence
 CC encodes a protein of the invention
 XX
 SQ Sequence 446 BP; 131 A; 90 C; 91 G; 134 T; 0 U; 0 Other;
 Query Match 3.0%; Score 18; DB 4; Length 446;
 Best Local Similarity 100.0%; Pred. No. 69;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 317 AACACAAATGGCTGAAA 334
 Db 157 AACACAAATGGCTGAAA 140
 RESULT 7
 AAS22506/C
 ID AAS22506 standard; cDNA; 471 BP.
 XX AC AAS22506;
 XX DT 24-OCT-2001 (first entry)
 XX DE Human cDNA encoding a novel human protein #72.
 XX KW Human; novel protein; ss; Antianaemic; osteopathic; antiinflammatory;
 XX KW immunomodulatory; cytostatic; neuroprotective; vulnery; nootropic;
 XX KW anticonvulsant; antiallergic; cerebroprotective; antifungal; antiviral;
 XX KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
 XX KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
 XX KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
 XX KW tissue regeneration; immune disorder.
 XX OS Homo sapiens.
 XX WO200155437-A2.
 XX PN 02-AUG-2001.
 XX PR 25-JAN-2001; 2001WO-US002623.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Tang YT, Liu C, Drmanac RT;
 XX WPI: 2001-451939/48.
 XX P-PSDB; AAU14201.
 XX PT Isolated polypeptides useful for treating anti-inflammatory diseases,
 XX nervous system disorders, and for regenerating bone and cartilage.
 XX PS Claim 1; Page 273-274; 894pp; English.
 XX
 CC The invention relates to polynucleotides encoding novel human proteins or
 CC their active domains. The polypeptides, polynucleotides and antibodies
 CC raised against the polypeptides are used in a method of treatment of a
 CC mammal and prevention of disorders caused by the aberrant protein
 CC expression or activity. The polypeptides can be used as molecular weight
 CC markers, food supplements, and in antibody production. The polypeptides
 CC are used to identify compounds which bind to the polypeptides.
 CC Polynucleotides of the invention are used as probes and primers, for

CC sequencing, for chromosome or gene mapping, in the production of
CC recombinant proteins, and in generating anti-sense DNA or RNA and in gene
CC therapy. Polypeptides of the invention can be used to target drugs to a
CC tumour, in assays to determine biological activity, to raise
CC antibodies/ elicit an immune response, to determine quantitative protein
CC levels, as tissue markers, and to isolate receptors or ligands.
CC Polypeptides of the invention may also be useful in treating platelet
CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
CC ligament and/or nerve tissue, wound healing, treating burns, promoting
CC the proliferation, differentiation and survival of stem cells, as a
CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
CC fungal infection or from autoimmunity, cancer, allergy, asthma, graft-
CC versus-host disease, eczema, haemophilia, thrombosis, anti-inflammatory
CC diseases, nervous system disorders, and infection. The present sequence
CC encodes a protein of the invention
XX
SQ Sequence 471 BP; 135 A; 96 C; 98 G; 142 T; 0 U; 0 Other;

Query Match 3.0%; Score 18; DB 4; Length 471;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 317 AACCAAAATGCGCTGAAA 334
Db 182 AACCAAAATGCGCTGAAA 165
|||||
|||

RESULT 8
ACH25483/c
ID ACH25483 standard; cDNA; 496 BP.
XX AC ACH25483;
XX
XX 13-OCT-2003 (first entry)
XX Human adult ovary cDNA #3863.
XX
XX Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
XX genome mapping; biodiversity; genetic disorder.
XX Homo sapiens.
XX US2003073623-A1.
XX 17-APR-2003.
XX 30-JUL-2001; 2001US-00918995.
XX 30-JUL-2001; 2001US-00918995.
XX (DRMA/) DRMANAC R T.
XX (LABA/) LABAT I.
XX (STAC/) STACHE-CRAIN B.
XX (DICK/) DICKSON M C.
XX (JONE/) JONES L W.
XX
XX Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX WPI; 2003-615964/58.
XX
XX New polynucleotide sequences obtained from various cDNA libraries, useful
XX as hybridization probes, as oligomers for PCR, for chromosome and gene
XX mapping, in the recombinant production of protein, or in generating
XX antisense DNA or RNA.
XX
XX Claim 1; SEQ ID NO 12695; 44pp; English.
XX
CC The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversity, or in identifying mutations

CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversity, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030073623
XX
SQ Sequence 496 BP; 138 A; 92 C; 84 G; 165 T; 0 U; 17 Other;

Query Match 3.0%; Score 18; DB 8; Length 496;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 82 ACAATGACAGCACTGACT 99
Db 446 ACAATGACAGCACTGACT 429
|||||
|||

RESULT 9
ACH28170/c
ID ACH28170 standard; cDNA; 498 BP.
XX AC ACH28170;
XX
XX 13-OCT-2003 (first entry)
XX Human adult ovary cDNA #6550.
XX
XX Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
XX genome mapping; biodiversity; genetic disorder.
XX Homo sapiens.
XX US2003073623-A1.
XX 17-APR-2003.
XX 30-JUL-2001; 2001US-00918995.
XX 30-JUL-2001; 2001US-00918995.
XX (DRMA/) DRMANAC R T.
XX (LABA/) LABAT I.
XX (STAC/) STACHE-CRAIN B.
XX (DICK/) DICKSON M C.
XX (JONE/) JONES L W.
XX
XX Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX WPI; 2003-615964/58.
XX
XX New polynucleotide sequences obtained from various cDNA libraries, useful
XX as hybridization probes, as oligomers for PCR, for chromosome and gene
XX mapping, in the recombinant production of protein, or in generating
XX antisense DNA or RNA.
XX
XX Claim 1; SEQ ID NO 15382; 44pp; English.
XX
CC The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversity, or in identifying mutations

CC responsible for genetic disorders and other traits. The nucleotide
 CC sequences are also useful as hybridisation probes, as oligomers for PCR,
 CC for chromosome and gene mapping, in the recombinant production of
 CC protein, or in generating antisense DNA or RNA. The purified polypeptide
 CC is useful for generating antibodies specific for it. The present sequence
 CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
 CC for this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from USPTO at
 CC seqdata.uspto.gov/sequence.html?DocID=20030073623

XX Sequence 498 BP; 141 A; 96 C; 105 G; 143 T; 0 U; 13 Other;

Query Match 3.0%; Score 18; DB 8; Length 498;
 Best Local Similarity 100.0%; Pred. No. 69;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 317 AACACAAATGGCCTGAAA 334
 DB 210 AACACAAATGGCCTGAAA 193

RESULT 10

ABN92747
 ID ABN92747 standard; DNA; 621 BP.

XX AC ABN92747;

XX DT 24-JUL-2002 (first entry)

XX DE Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:2210.

XX KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
 XX antibiotic; gene therapy; gene; ds.

XX OS Staphylococcus epidermidis.

XX PN US6380370-B1.

XX PD 30-APR-2002.

XX PF 13-AUG-1998; 98US-00134001.

XX PR 14-AUG-1997; 97US-0055779P.

XX PR 08-NOV-1997; 97US-0064964P.

XX PA (GENO-) GENOME THERAPEUTICS CORP.

XX PI Doucette-Stamm LA, Bush D;

XX DR WPI; 2002-381255/41.

XX DR P-PSDB; ABP40202.

XX PT Novel isolated nucleic acid encoding a Staphylococcus epidermis

XX PT polypeptide, useful for diagnosing and treating bacterial infections.

XX PS Disclosure; SEQ ID NO 2210; 267pp; English.

XX CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
 CC given in ABP3124 to ABP37960. The S. epidermidis sequences have
 CC antibacterial activity and can be used in gene therapy. The sequences can
 CC also be used in the diagnosis and treatment of bacterial infections,
 CC particularly S. epidermidis infections. The sequences can be used to
 CC screen for compounds able to interfere with the S. epidermidis life cycle
 CC or inhibit S. epidermidis infection. N.B. The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from the USPTO web site

XX Sequence 621 BP; 234 A; 94 C; 102 G; 191 T; 0 U; 0 Other;

Query Match 3.0%; Score 18; DB 6; Length 621;
 Best Local Similarity 100.0%; Pred. No. 69;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 247 TATTATTTCAGATTAT 264
 DB 238 TATTATTTCAGATTAT 255

RESULT 11

ABL26531
 ID ABL26531 standard; DNA; 1125 BP.

XX AC ABL26531;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 31066.

XX KW Drosophila; developmental biology; cell signalling; insecticide;
 XX pharmaceutical; gene; ds.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US009231.

XX PR 23-MAR-2000; 2000US-0191637P.

XX PR 11-JUL-2000; 2000US-00614150.

XX PA (PEKE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
 XX genes from Drosophila and for elucidating cell signalling and cell-cell
 XX interactions.

XX PS Claim 1; SEQ ID NO 31066; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent
 XX capable of detecting 1000 or more genes from Drosophila. The invention is
 XX useful in developmental biology and in elucidating cell signalling and
 XX cell-cell interactions in higher eukaryotes for the development of
 XX insecticides, therapeutics and pharmaceutical drugs. The invention
 XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 XX sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
 XX ABB72072). The sequence data for this patent did not form part of the
 XX printed specification, but was obtained in electronic format directly
 XX from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 1125 BP; 304 A; 251 C; 291 G; 279 T; 0 U; 0 Other;

Query Match 3.0%; Score 18; DB 4; Length 1125;
 Best Local Similarity 100.0%; Pred. No. 68;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 583 GAATCAGCGGATACCAT 600

DB 826 GAATCAGCGGATACCAT 843

RESULT 12

ADA00895/c
 ID ADA00895 standard; cDNA; 2997 BP.

XX AC ADA00895;

XX DT 06-NOV-2003 (first entry)

XX DE Mouse osteoclast related protein Gl encoding cDNA SEQ ID NO:1.

XX GI, bone resorption; osteoclast; osteopathic; gene therapy;
KW bone related disease; osteoporosis; osteoporosis; osteoclast maturation;
KW osteoclast related protein; gene; ss.
XX Mus musculus.
FH Key Location/Qualifiers
FT CDS 45..1061
FT /*tag= a
FT /product= "osteoclast related protein GI"
XX WO2003029283-A2.
XX
XX 10-APR-2003.
XX
XX 25-SEP-2002; 2002WO-EP010721.
XX
XX 27-SEP-2001; 2001CA-02357987.
XX (AVET) AVENTIS PHARMA SA.
XX Vacher J, Chalhoub N, Benachenou N;
XX
XX WPI; 2003-381602/36.
XX P-PSDB; ADA00896.
XX New GI polypeptides and genes encoding them, useful for preventing or
PT treating bone related diseases e.g. osteoporosis or osteoporosis, and
PT for modulating proper osteoclast maturation and bone tissue resorption.
XX
XX Claim 6; Fig 8; 76pp; English.
XX The present invention describes an isolated or purified nucleic acid
CC molecule (I) encoding a mammalian GI polypeptide which modulates bone
CC resorption in osteoclast cells. (I) has osteopathic activity and can be
CC used in gene therapy. GI genes and polypeptides are useful for preventing
CC or treating bone related diseases such as osteoporosis or osteoporosis,
CC and for proper osteoclast maturation and bone tissue resorption. Probes
CC and antibodies raised against the GI gene product can be used in
CC hybridisation and immunological assays to screen for and detect the
CC presence of either a normal or mutated gene or gene product. The present
CC sequence encodes the mouse GI osteoclast related protein, from the
CC present invention.
XX
XX Sequence 2997 BP; 818 A; 642 C; 696 G; 841 T; 0 U; 0 Other;
SQ
Query Match 3.0%; Score 18; DB 7; Length 2997;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 77 CAGGAACAAATGACAGCAC 94
Db 2587 CAGGAACAAATGACAGCAC 2570
RESULT 13
ABL26530
ID ABL26530 standard; DNA; 3125 BP.
AC ABL26530;
XX
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 31063.
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmacological; gene; ds.
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX

PD 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
XX Claim 1; SEQ ID NO 31063; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 3125 BP; 966 A; 597 C; 649 G; 913 T; 0 U; 0 Other;
SQ
Query Match 3.0%; Score 18; DB 4; Length 3125;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 583 CAATCAGCGGATACCACT 600
Db 1826 CAATCAGCGGATACCACT 1843
RESULT 14
ABL21028/c
ID ABL21028 standard; DNA; 11920 BP.
XX
XX ABL21028;
XX
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 14557.
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmacological; gene; ds.
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.

```
PT interactions.
XX
PS Claim 1; SEQ ID NO 14557; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL161176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL161175) and the encoded proteins (ABBS57737-
CC ABB2072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 11920 BP; 3521 A; 2455 C; 2509 G; 3435 T; 0 U; 0 Other;
Query Match 3.0%; Score 18; DB 4; Length 11920;
Best Local Similarity 100.0%; Pred.No. 67;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 373 GAGGAAAGCCAGTTT 390
Db 2140 GAGGAAAGCCAGTTT 2123
RESULT 15
ABN35170/c
ID ABN35170 standard; DNA; 60 BP.
XX
AC ABN35170;
XX
DI 15-JUL-2002 (first entry)
XX
DE Human spliced transcript detection oligonucleotide SEQ ID NO:7918.
XX
KW Human; mouse; rat; splice transcript; detection; RNA transcript;
KW splice variant; transcriptome; oligonucleotide library; ss.
XX
OS Homo sapiens.
XX
PN WO200210449-A2.
XX
PD 07-FEB-2002.
XX
PF 20-JUL-2001; 2001WO-IB001503.
XX
PR 28-JUL-2000; 2000US-0221607P.
PR 02-MAY-2001; 2001US-0287724P.
XX
PA (COMP-) COMPUEN INC.
XX
PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
XX
DR WPI; 2002-257383/30.
XX
PT New oligonucleotide libraries comprising oligonucleotides which
PT selectively hybridize to mRNAs transcribed from a transcription unit of a
PT genome, useful for detecting tissue-, pathology-, and developmental-
PT specific genes.
XX
PS Example 1; SEQ ID NO 7918; 47pp; English.
XX
CC The present invention describes oligonucleotide libraries for detecting
CC messenger RNAs that populate a (sub-)transcriptome, where the (sub-)
CC transcriptome comprises messenger RNAs transcribed from multiple
CC transcription units that populate a genome. The library comprises several
CC oligonucleotides, each capable of hybridising selectively to a set of
CC messenger RNAs transcribed from a given transcription unit of the genome,
CC which encodes one or more messenger RNA splice variants. The
CC oligonucleotide libraries are useful for detecting mRNAs from a
CC biological sample, in expression profiling studies, in qualitatively or
CC quantitatively characterising the corresponding transcriptome, and in
```

```
CC detecting RNA transcripts and splice variants of human or animal
CC transcriptomes. The libraries may also be used as specialised mini
CC libraries to detect transcripts of a sub-transcriptome under a particular
CC biological or pathological state, and so allowing the detection of tissue
CC - and pathology-specific genes such as those genes only expressed in
CC specific tissue under a specific pathological condition; to detect
CC developmental specific genes; and to detect RNA transcripts and splice
CC variants of a transcriptome of a patient suffering from a particular
CC disorder. ABN27253 to ABN59589 represent oligonucleotide sequences from
CC rats, humans and mice, which are used in the exemplification of the
CC present invention. N.B. The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 60 BP; 16 A; 15 C; 15 G; 14 T; 0 U; 0 Other;
Query Match 2.8%; Score 17; DB 6; Length 60;
Best Local Similarity 100.0%; Pred.No. 2.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 373 GAGGAAAGCCAGTTT 389
Db 59 GAGGAAAGCCAGTTT 43
Search completed: July 30, 2004, 15:35:53
Job time : 457 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 30, 2004, 15:19:50 ; Search time 99 Seconds
(without alignments)
3413.789 Million cell updates/sec

Title: US-09-868-338-7_COPY_1117_1725
Perfect score: 609
Sequence: 1 atgataagaatcaatgacct.....cggataccattatcaaaccta 609

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTCUS COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	19	3.1	855	4	US-09-328-352-340
2	18	3.0	621	4	US-09-134-001C-2210
3	17	2.8	192	4	US-09-107-532A-6
4	17	2.8	700	3	US-08-718-904-109
5	17	2.8	700	4	US-09-449-249-109
6	17	2.8	711	4	US-09-107-532A-2667
7	17	2.8	858	4	US-09-107-532A-2722
8	17	2.8	1127	2	US-07-735-068-1
9	17	2.8	1497	4	US-09-107-532A-3337
10	17	2.8	2351	4	US-09-620-312B-848
11	17	2.8	2531	4	US-09-468-656A-5
12	17	2.8	2561	4	US-09-347-878-25
13	17	2.8	8195	4	US-08-961-527-94
14	16	2.6	183	4	US-09-134-001C-1676
15	16	2.6	255	3	US-08-821-994-58
16	16	2.6	285	4	US-09-312-283C-411
17	16	2.6	292	4	US-09-312-283C-411
18	16	2.6	339	4	US-09-312-283C-7400
19	16	2.6	456	4	US-09-976-594-1004
20	16	2.6	502	4	US-09-166-350-2
21	16	2.6	523	4	US-09-833-381-1256
22	16	2.6	527	3	US-09-188-930-118
23	16	2.6	527	4	US-09-312-283C-118
24	16	2.6	527	4	US-09-312-283C-118
25	16	2.6	568	4	US-09-621-976-472
26	16	2.6	585	4	US-09-312-283C-371
27	16	2.6	653	4	US-09-833-381-1257

28	16	2.6	708	4	US-09-833-381-1258	Sequence 1258, Ap
29	16	2.6	735	4	US-09-833-381-48	Sequence 48, Appl
30	16	2.6	774	4	US-09-134-000C-520	Sequence 520, Appl
31	16	2.6	784	4	US-09-833-381-1255	Sequence 1255, Ap
32	16	2.6	804	4	US-09-252-991A-6060	Sequence 6060, Ap
33	16	2.6	846	4	US-09-543-681A-629	Sequence 629, Appl
34	16	2.6	891	4	US-09-134-001C-2153	Sequence 2153, Ap
35	16	2.6	943	4	US-09-833-381-1359	Sequence 1259, Ap
36	16	2.6	952	4	US-09-833-381-1260	Sequence 1260, Ap
37	16	2.6	1035	4	US-09-540-236-782	Sequence 782, Appl
38	16	2.6	1068	3	US-09-211-542A-11	Sequence 11, Appl
39	16	2.6	1119	1	US-07-854-596B-30	Sequence 30, Appl
40	16	2.6	1122	4	US-09-374-038-13	Sequence 13, Appl
41	16	2.6	1122	4	US-09-658-179-13	Sequence 13, Appl
42	16	2.6	1158	4	US-09-374-038-14	Sequence 14, Appl
43	16	2.6	1158	4	US-09-658-179-14	Sequence 14, Appl
44	16	2.6	1209	4	US-09-374-038-11	Sequence 11, Appl
45	16	2.6	1209	4	US-09-658-179-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-09-328-352-340
; Sequence 340, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 340
; LENGTH: 855
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-340

Query Match 3.1%; Score 19; DB 4; Length 855;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	115	GGAAATCGACTTTGCTCA	133
DB	199	GGAAATCGACTTTGCTCA	217

RESULT 2
US-09-134-001C-2210
; Sequence 2210, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2210
; LENGTH: 621
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2210

Query Match 3.0%; Score 18; DB 4; Length 621;

```

; Patent No. 6037329
; GENERAL INFORMATION:
; APPLICANT: Baird, J. Andrew
; APPLICANT: Chandler, Lois Ann
; APPLICANT: Sosnowski, Barbara A.
; TITLE OF INVENTION: COMPOSITIONS CONTAINING NUCLEIC ACIDS AND LIGANDS FOR THERAPE
; NUMBER OF SEQUENCES: 128
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/718,904
; FILING DATE: 24-SEP-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: NO. 6037329tenburg Ph.D., Carol
; REGISTRATION NUMBER: 39,317
; REFERENCE/DOCKET NUMBER: 760100.415C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-718-904-109

Query Match 2.8%; Score 17; DB 3; Length 700;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 322 AAATGCCCTGAAATTC 338
Db 314 AAATGCCCTGAAATTC 330

RESULT 5
US-09-449-249-109
; Sequence 109, Application US/09449249
; Patent No. 6503886
; GENERAL INFORMATION:
; APPLICANT: Baird, J. Andrew
; Sosnowski, Barbara A.
; TITLE OF INVENTION: COMPOSITIONS CONTAINING NUCLEIC ACIDS AND LIGANDS FOR THERAI
; NUMBER OF SEQUENCES: 128
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/449,249
; FILING DATE: 24-No. 6503886-1999
; CLASSIFICATION: <Unknown>

```

```

; Patent No. 6037329
; GENERAL INFORMATION:
; APPLICANT: Baird, J. Andrew
; APPLICANT: Chandler, Lois Ann
; APPLICANT: Sosnowski, Barbara A.
; TITLE OF INVENTION: COMPOSITIONS CONTAINING NUCLEIC ACIDS AND LIGANDS FOR THERAPE
; NUMBER OF SEQUENCES: 128
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/718,904
; FILING DATE: 24-SEP-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: NO. 6037329tenburg Ph.D., Carol
; REGISTRATION NUMBER: 39,317
; REFERENCE/DOCKET NUMBER: 760100.415C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-718-904-109

Query Match 2.8%; Score 17; DB 3; Length 700;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 322 AAATGCCCTGAAATTC 338
Db 314 AAATGCCCTGAAATTC 330

RESULT 5
US-09-449-249-109
; Sequence 109, Application US/09449249
; Patent No. 6503886
; GENERAL INFORMATION:
; APPLICANT: Baird, J. Andrew
; Sosnowski, Barbara A.
; TITLE OF INVENTION: COMPOSITIONS CONTAINING NUCLEIC ACIDS AND LIGANDS FOR THERAI
; NUMBER OF SEQUENCES: 128
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/449,249
; FILING DATE: 24-No. 6503886-1999
; CLASSIFICATION: <Unknown>

```

```
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/718,904
; FILING DATE: 24-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 650388stemburg Ph.D., Carol
; REGISTRATION NUMBER: 39,317
; REFERENCE/DOCKET NUMBER: 760100.415C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 109:
US-09-449-249-109

Query Match      2.8%; Score 17; DB 4; Length 700;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

;
; RESULT 6
; US-09-107-532A-2667
; Sequence 2667, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-8277
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 2667:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 711 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; INFORMATION FOR SEQ ID NO: 2667:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 711 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO

;
; US-09-107-532A-2667
; Sequence 2667, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-8277
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 2667:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 711 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO

;
; US-09-107-532A-2722
; Sequence 2722, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-8277
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 2722:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 858 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...858
; SEQUENCE DESCRIPTION: SEQ ID NO: 2722:
US-09-107-532A-2722

Query Match      2.8%; Score 17; DB 4; Length 858;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 109 GGTTCAGGAATCGAC 125
DB 145 GGTTCAGGAATCGAC 161

RESULT 8

US-07-735-068-1
; Sequence 1, Application US/07735068
; GENERAL INFORMATION:
; APPLICANT: Kumaz, Chanakanti C.
; TITLE OF INVENTION: No. 5885769el Screening Systems
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schering-Plough Corporation
; STREET: One Giralda Farms
; CITY: Madison
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07940-1000
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.5
; SOFTWARE: Microsoft Word 4.00B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07735.068
; FILING DATE: 19910724
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US07/655966
; FILING DATE: 14-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Dulak, No. 5885769man C.
; REGISTRATION NUMBER: 31,608
; REFERENCE/DOCKET NUMBER: ID0178K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 822 7375
; TELEFAX: 201 822 7039
; TELEX: 219165
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1127 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
US-07-735-068-1

Query Match 2.8%; Score 17; DB 2; Length 1127;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 322 AAATGGCGCTGAATTC 338
DB 530 AAATGGCGCTGAATTC 546

RESULT 9

US-09-107-532A-3337
; Sequence 3337, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA

ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Atinello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3337:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1497 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...1497
; SEQUENCE DESCRIPTION: SEQ ID NO: 3337:
US-09-107-532A-3337

Query Match 2.8%; Score 17; DB 4; Length 1497;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 AAATCGACTTTGCTCAA 134
DB 148 AAATCGACTTTGCTCAA 164

RESULT 10

US-09-620-312D-848/c
; Sequence 848, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yuning
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D

; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pc_FU_genes Version 1.0
; SEQ ID NO 848
; LENGTH: 2351
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (46)..(1422)
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(2351)
; OTHER INFORMATION: n = a,t,c or g
US-09-620-312D-848

Query Match 2.8%; Score 17; DB 4; Length 2351;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 374 AGGAAAGCCAGTTT 390
DB 1981 AGGAAAGCCAGTTT 1965

RESULT 11
US-09-468-656A-5
; Sequence 5, Application US/09468656A
; Patent No. 6582706
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-444
; CURRENT APPLICATION NUMBER: US/09/468,656A
; CURRENT FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 2531
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-468-656A-5

Query Match 2.8%; Score 17; DB 4; Length 2531;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 329 CTGAAATTCCTCAAGTA 345
DB 2294 CTGAAATTCCTCAAGTA 2310

RESULT 12
US-09-347-878-25
; Sequence 25, Application US/09347878C
; Patent No. 6376210
; GENERAL INFORMATION:
; APPLICANT: Yuan, Chong
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ASSAYING ANALYTES
; FILE REFERENCE: 25885-1651
; CURRENT APPLICATION NUMBER: US/09/347,878C
; CURRENT FILING DATE: 1999-07-06
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 25
; LENGTH: 2561
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1131)..(2399)
; OTHER INFORMATION: Escherichia coli nucleic acid encoding
; OTHER INFORMATION: f0lypolyglutamate synthetase-dihydrofolate
; OTHER INFORMATION: synthetase
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: M32445/GenBank
US-09-347-878-25

Query Match 2.8%; Score 17; DB 4; Length 2561;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 583 GAATCAGCGATACCAT 599
DB 1116 GAATCAGCGATACCAT 1132

RESULT 13
US-08-961-527-94
; Sequence 94, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8195 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-94

Query Match 2.8%; Score 17; DB 4; Length 8195;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 329 CTGAAATTCCTCAAGTA 345
DB 7895 CTGAAATTCCTCAAGTA 7911

RESULT 14
US-09-134-001C-1676
; Sequence 1676, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-09
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1676
; LENGTH: 183
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1676

Query Match 2.6%; Score 16; DB 4; Length 183;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 480 AGCTCTAGATTAAACA 495
Db 123 AGCTCTAGATTAAACA 138

RESULT 15
US-08-821-994-58/c
; Sequence 58, Application US/08821994A
; Patent No. 6228643
; GENERAL INFORMATION:
; APPLICANT: Greenland, Andrew J
; APPLICANT: Thomas, Didier RP
; APPLICANT: Jepson, Ian
; TITLE OF INVENTION: Promoters
; FILE REFERENCE: PPD 50108
; CURRENT APPLICATION NUMBER: US/08/821,994A
; CURRENT FILING DATE: 1997-03-22
; EARLIER APPLICATION NUMBER: PCT/GB97/00729
; EARLIER FILING DATE: 1997-03-18
; EARLIER APPLICATION NUMBER: GB 9606062.9
; EARLIER FILING DATE: 1996-03-22
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 58
; LENGTH: 255
; TYPE: DNA
; ORGANISM: Brassica napus
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 136, 187, 192
; OTHER INFORMATION: n is unknown
US-08-821-994-58

Query Match 2.6%; Score 16; DB 3; Length 255;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 547 ACCGTTGTTGCTA 562
Db 216 ACCGTTGTTGCTA 201

Search completed: July 30, 2004, 17:25:10
Job time : 100 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 30, 2004, 15:13:30 ; Search time 2790 Seconds
(without alignments)

Sequence: 1 atgataagaatcaatgacct.....cggataccattatcaacta 609
6518.304 Million cell updates/sec

Title: US-09-868-338-7_COPY_1117_1725

Perfect score: 609

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues
Word size : 0

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: listing first 45 summaries

Database :

```

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmd:*
5: em_estcov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_man:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gssi:*
29: gb_gss2:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Match	Length	DB	ID	Description
C 1	20	3.3	231	9	AV279945	AV279945	AV279945
C 2	20	3.3	244	9	AV283498	AV283498	AV283498
C 3	20	3.3	303	10	BS522412	BS522412	BS522412
C 4	20	3.3	353	10	AW319130	AW319130	AW319130

5	20	3.3	362	10	AW210581	AW210581	AW210581
6	20	3.3	368	9	AT573872	AT573872	AT573872
7	20	3.3	379	9	AT874876	AT874876	AT874876
8	20	3.3	383	9	AW012564	AW012564	AW012564
9	20	3.3	392	9	AT317342	AT317342	AT317342
10	20	3.3	417	9	AA397040	AA397040	AA397040
11	20	3.3	427	9	AA260497	AA260497	AA260497
12	20	3.3	440	13	EX635140	EX635140	EX635140
13	20	3.3	445	10	BB690607	BB690607	BB690607
14	20	3.3	455	9	AI427592	AI427592	AI427592
15	20	3.3	463	10	BB747871	BB747871	BB747871
16	20	3.3	464	9	AA250581	AA250581	AA250581
17	20	3.3	471	10	BB748528	BB748528	BB748528
18	20	3.3	478	13	BQ621602	BQ621602	BQ621602
19	20	3.3	481	9	AA274854	AA274854	AA274854
20	20	3.3	504	9	AA112895	AA112895	AA112895
21	20	3.3	518	9	AI182000	AI182000	AI182000
22	20	3.3	569	9	AI196816	AI196816	AI196816
23	20	3.3	600	9	AI042942	AI042942	AI042942
24	20	3.3	611	9	AI035653	AI035653	AI035653
25	20	3.3	748	12	BG702850	BG702850	BG702850
26	20	3.3	754	9	AI326387	AI326387	AI326387
27	20	3.3	774	9	AI046692	AI046692	AI046692
28	20	3.3	804	9	AI255825	AI255825	AI255825
29	19	3.1	249	10	BB602955	BB602955	BB602955
30	19	3.1	306	10	BB217097	BB217097	BB217097
31	19	3.1	311	9	AL823190	AL823190	AL823190
32	19	3.1	419	10	BF659278	BF659278	BF659278
33	19	3.1	522	29	EX124852	EX124852	EX124852
34	19	3.1	528	13	BQ767175	BQ767175	BQ767175
35	19	3.1	544	13	BQ444036	BQ444036	BQ444036
36	19	3.1	544	14	CD564335	CD564335	CD564335
37	19	3.1	552	28	AQ873948	AQ873948	AQ873948
38	19	3.1	567	28	BH350329	BH350329	BH350329
39	19	3.1	580	14	CA533799	CA533799	CA533799
40	19	3.1	585	28	BZ694104	BZ694104	BZ694104
41	19	3.1	596	29	CG807230	CG807230	CG807230
42	19	3.1	600	12	BI988460	BI988460	BI988460
43	19	3.1	600	13	BU702558	BU702558	BU702558
44	19	3.1	604	14	CF257500	CF257500	CF257500
45	19	3.1	629	14	CD562070	CD562070	CD562070

ALIGNMENTS

AV279945 231 bp mRNA linear EST 05-NOV-1999
AV279945 RIKEN full-length enriched, adult male testis (DRI08) Mus
musculus cDNA clone 493411H02 3' similar to A0011080 Mus musculus
mRNA for alpha-albumin protein, mRNA sequence.

AV279945
AV279945.1 GI:6267982

Mus musculus (house mouse)

EST

KEYWORDS

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases) 1 to 231
Konno.H., Aizawa.K., Akahira.S., Akiyama.J., Carninci.P., Endo.T.,
Fukuda.S., Fukunishi.Y., Harada.A., Hayatsu.N., Hirozane.T., Hori.F.,
Ishii.Y., Ishikawa.T., Itoh.M., Izawa.M., Kadota.K., Kagawa.I.,
Kai.C., Kawai.J., Kikuchi.N., Kojima.Y., Koya.S., Kusakabe.M.,
Matsuyama.T., Miki.R., Mizuno.Y., Nakamura.M., Oda.H., Okazaki.Y.,
Owa.C., Ozawa.Y., Saito.H., Sano.M., Sato.K., Shibata.K.,
Shibata.Y., Shigemoto.Y., Shiraki.T., Sogabe.Y., Sugahara.Y.,
Suzuki.H., Suzuki.H., Takahashi.F., Tateo.M., Tomimaga.N.,
Tsukuda.Y., Watanabe.S., Watanabe.S., Yamamura.T., Yasunishi.A.,
Yoshida.T., Yoshikawa.A., Yoshino.M., Muramatsu.M. and Hayashizaki.Y.
RIKEN Mouse ESTs (Konno.H., et al. 1999)
Unpublished (1999)
Contact: Yoshihide Hayashizaki

was primed with an oligo(dT) primer
[ATGGGCGCTTTTCTTTTCTTTTCTTTT]; double-stranded cDNA was
ligated to a DralII adaptor [GTTGGCCACTG], digested
and cloned into distinct DraIII sites of the pME18S-FL3
vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
be used to isolate the cDNA insert. Size selection was
performed to exclude fragments <1.5kb. Library
constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTCTGCTCTAAAGCTGG and 3' end
primer CCACCTGAGCTGAGCACA."

ORIGIN

Query Match 3.3%; Score 20; DB 10; Length 353;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 382 CCAGTTTGAACCTCTCTGG 401
|||||
Db 163 CCAGTTTGAACCTCTCTGG 182

RESULT 5

AW210581 362 bp mRNA linear EST 03-DEC-1999
LOCUS um62a11.x1 Sugano mouse kidney mkia Mus musculus cDNA clone
DEFINITION IMAGE:2285948 3' similar to SM:AFAM_MOUSE 089020 AFAMIN PRECURSOR
; mRNA sequence.

ACCESSION AW210581.1 GI:6516521

VERSION AW210581

KEYWORDS EST.

SOURCE Mus musculus

ORGANISM Mus musculus (house mouse)

REFERENCE

AUTHORS

Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,
Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,
Ritter, E., Kohn, S., Shin, T., Jackson, J., Cardenas, M., McCann, R.,
Waterston, R. and Wilson, R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)

TITLE

JOURNAL

COMMENT

Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:1008544

Seq primer: custom primer used.

FEATURES

source

1. .362
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:2285948"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="Sugano mouse kidney mkia"

/note="Organ: kidney; Vector: pME18S-FL3; Site 1: DraIII

(CACTGTGTG); Site 2: DraIII (CACCATGTG); 1st strand cDNA

was primed with an oligo(dT) primer

[ATGGGCGCTTTTCTTTTCTTTTCTTTT]; double-stranded cDNA was

ligated to a DraIII adaptor [GTTGGCCACTG], digested

and cloned into distinct DraIII sites of the pME18S-FL3

vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should

be used to isolate the cDNA insert. Size selection was

performed to exclude fragments <1.5kb. Library

constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTCTGCTCTAAAGCTGG and 3' end
primer CCACCTGAGCTGAGCACA."

ORIGIN

Query Match 3.3%; Score 20; DB 10; Length 362;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 382 CCAGTTTGAACCTCTCTGG 401
|||||
Db 168 CCAGTTTGAACCTCTCTGG 187

RESULT 6

AW210581

LOCUS

DEFINITION

IMAGE:1908299 3' similar to SM:AFAM_RAT P36953 AFAMIN PRECURSOR ;

mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Marra M/WashU-NCI Mouse EST Project 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:976495

Seq primer: custom primer used.

FEATURES

source

1. .368

Location/Qualifiers

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL"

/db_xref="taxon:10090"

/clone="IMAGE:1908299"

/sex="female"

/dev_stage="adult"

/lab_host="DH10B"

/clone_lib="Sugano mouse kidney mkia"

/note="Organ: kidney; Vector: pME18S-FL3; Site 1: DraIII

(CACTGTGTG); Site 2: DraIII (CACCATGTG); 1st strand cDNA

was primed with an oligo(dT) primer

[ATGGGCGCTTTTCTTTTCTTTTCTTTT]; double-stranded cDNA was

ligated to a DraIII adaptor [GTTGGCCACTG], digested

and cloned into distinct DraIII sites of the pME18S-FL3

vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should

be used to isolate the cDNA insert. Size selection was

performed to exclude fragments <1.5kb. Library

constructed by Dr. Sumio Sugano (University of Tokyo

Institute of Medical Science). Custom primers for

sequencing: 5' end primer CTCTGCTCTAAAGCTGG and 3' end

primer CCACCTGAGCTGAGCACA."

ORIGIN

Query Match 3.3%; Score 20; DB 9; Length 368;

[illegible]

DEFINITION uj13e04.x1 Sugano mouse kidney mkia Mus musculus cDNA clone IMAGE:1907934 3' similar to SW:AFAM_RAT P36953 AFAMIN PRECURSOR ; mRNA sequence.

ACCESSION AI317342

VERSION AI317342.1 GI:4032609

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 392)

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE The WashU-HHMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:976130

Seq primer: custom primer used.

FEATURES source Location/Qualifiers 1..392 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL" /db_xref="taxon:10090" /clone="IMAGE:1907934" /sex="female" /dev_stage="adult" /lab_host="PH10B" /clone_lib="Sugano mouse kidney mkia" /notes="Organ: kidney; Vector: pME18S-FL3; Site 1: DralII (CACTGTGTG); Site 2: DraIII (CACCATGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTCCTTTTGTATTTTCTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTGGCTTACTGGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTCTGCTCTTAAAGTGGG and 3' end primer CGACCTGCAGCTCGACACA."

ORIGIN Query Match 3.3%; Score 20; DB 9; Length 392; Best Local Similarity 100.0%; Pred. No. 33; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 382 CCAGTTTTTGAACCTCTCTGG 401

Db 198 CCAGTTTTTGAACCTCTCTGG 217

RESULT 10

AA397040/c

LOCUS mx84h09.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:1907934 5'

DEFINITION similar to SW:AFAM_RAT P36953 AFAMIN PRECURSOR ; mRNA sequence.

ACCESSION AA397040

VERSION AA397040.1 GI:2050141

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 417)

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE The WashU-HHMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:426649

High quality sequence stop: 393.

FEATURES source Location/Qualifiers 1..417 /organism="Mus musculus" /mol_type="mRNA" /db_xref="taxon:10090" /clone="IMAGE:693089" /tissue_type="Liver" /lab_host="DH10B" /clone_lib="Soares mouse NML" /notes="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I oligo(dT) primer [5' TGTACCAATCATGAGTGGGCGGCGGCAATCTTTTCTTTTCTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."

ORIGIN Query Match 3.3%; Score 20; DB 9; Length 417; Best Local Similarity 100.0%; Pred. No. 33; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 382 CCAGTTTTTGAACCTCTCTGG 401

Db 346 CCAGTTTTTGAACCTCTCTGG 327

RESULT 11

AA260497/c

LOCUS va59i0.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:747234 5'

DEFINITION similar to SW:AFAM_RAT P36953 AFAMIN PRECURSOR ; mRNA sequence.

ACCESSION AA260497

VERSION AA260497.1 GI:1896999

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 427)

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE The WashU-HHMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mousest@watson.wustl.edu
This clone is available royalty-free through LLM ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:456218
High quality sequence stop: 406.
Location/Qualifiers
1. 427
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:747234"
/tissue_type="Liver"
/lab_host="DH10B"
/clone_lib="Soares mouse NML"
/note="vector: p773D-Fac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGAGCGCGCAATCTTTTCTTTT 3'];
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p773 vector. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."

FEATURES
source
Query Match 3.3%; Score 20; DB 9; Length 427;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 382 CCAGTTTGAACCTCTGG 401
Db 298 CCAGTTTGAACCTCTGG 279

ORIGIN
Mus musculus (house mouse)
BX635140 pBluescript Lion Mus musculus cDNA clone LIONp462D02381
EST 12-AUG-2003
LOCUS
DEFINITION
BX635140 440 bp mRNA linear EST 12-AUG-2003
3', mRNA sequence.
ACCESSION
BX635140
VERSION
BX635140.1 GI:33615015
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 440)
REFERENCE
Henrich,J., Hermanns,J., Kranz,H., Loebbert,R., Schluter,T.,
Schuette,D., Weindel,M., Heil,O., Ebert,L., Neubert,P., Peters,M.,
Radelof,U., Schneider,D. and Korn,B.
Mouse ArrayTAG cDNA (LION)
Unpublished (2003)
Contact: Ina Rolfes
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
RZPD; LIONp462D02381.
RZPDLib;
Mouse ArrayTAG cDNA (LION)
http://www.rzpd.de/cgi-bin/products/showlib.pl.cgi?response?libNo=4
62 Contact: Ina Rolfes
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heuberweg 6, D-14059 Berlin, Germany
Tel: +49 30 32639 101
Fax: +49 30 32639 111
www.rzpd.de
This clone is available royalty-free from RZPD;
contact RZPD (clone@rzpd.de) for further information. Seq primer:
RP: CAGGAACAGCTATGAC.
Location/Qualifiers

source
1. 440
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="LIONp462D02381"
/lab_host="DH10B"
/clone_lib="pBluescript Lion"

ORIGIN
Query Match 3.3%; Score 20; DB 13; Length 440;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 382 CCAGTTTGAACCTCTGG 401
Db 189 CCAGTTTGAACCTCTGG 208

RESULT 13
BB690607/c
LOCUS
DEFINITION
BB690607 RIKEN full-length enriched, 12 days embryo female
mullerian duct Mus musculus cDNA clone 6820448P03 3', mRNA
sequence.
BB690607
BB690607.1 GI:16017340
EST.
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 445)
REFERENCE
Akimura,T., Arakawa,T., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K.,
Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
Hayatsu,N., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R.,
Matsuyama,T., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,
Okazaki,Y., Okido,T., Saito,R., Shinagawa,A., Shiraki,T.,
Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Takahashi,F., Takaku-Akaihi,S.,
Sogabe,Y., Suzuki,H., Tagawa,A., Tanaka,A., Tanaka,T.,
Tanaka,T., Tomaru,A., Toya,T., Watabiki,A., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp/
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.

FEATURES
source
1. 440
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="LIONp462D02381"
/lab_host="DH10B"
/clone_lib="pBluescript Lion"

ORIGIN
Query Match 3.3%; Score 20; DB 13; Length 440;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 382 CCAGTTTGAACCTCTGG 401
Db 189 CCAGTTTGAACCTCTGG 208

RESULT 13
BB690607/c
LOCUS
DEFINITION
BB690607 RIKEN full-length enriched, 12 days embryo female
mullerian duct Mus musculus cDNA clone 6820448P03 3', mRNA
sequence.
BB690607
BB690607.1 GI:16017340
EST.
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 445)
REFERENCE
Akimura,T., Arakawa,T., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K.,
Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
Hayatsu,N., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R.,
Matsuyama,T., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,
Okazaki,Y., Okido,T., Saito,R., Shinagawa,A., Shiraki,T.,
Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Takahashi,F., Takaku-Akaihi,S.,
Sogabe,Y., Suzuki,H., Tagawa,A., Tanaka,A., Tanaka,T.,
Tanaka,T., Tomaru,A., Toya,T., Watabiki,A., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp/
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.

Search completed: July 30, 2004, 17:23:28
Job time : 2797 secs

This Page Blank (uspig)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 30, 2004, 07:14:14 ; Search time 3631 Seconds
(without alignments)
7269.597 Million cell updates/sec

Title: US-09-868-338-7_COPY_1117_1725

Perfect score: 609

Sequence: 1 atgatgaaatcaatgacct.....cggataccattatcaaaacta 609

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hgt.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rod.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_hngo_hum.*

40: em_hngo_mus.*

41: em_hngo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	96.4	15.8	11964	1	AE010640	AE010640 Fusobacte
C 2	91.8	15.1	341553	1	BX248355	BX248355 Corynebact
C 3	84.6	13.9	347961	1	BX571657	BX571657 Wolinella
C 4	83.8	13.8	54310	1	AP003515	AP003515 Clostridi
C 5	83.6	13.7	301439	1	AE015943	AE015943 Clostridi
C 6	83	13.6	10575	1	AE000977	AE000977 Archaeogl
C 7	82.4	13.5	10415	1	AE007598	AE007598 Clostridi
C 8	82	13.5	12929	1	AE013394	AE013394 Methanosa
C 9	80.8	13.3	96208	1	AL596174	AL596174 Listeria
C 10	80.8	13.3	311208	6	AX417049	AX417049 Sequence
C 11	80.8	13.3	349980	6	AX413018	AX413018 Sequence
C 12	80.8	13.3	349980	6	AX417036	AX417036 Sequence
C 13	80.6	13.2	11551	1	AE006342	AE006342 Lactococc
C 14	80.6	13.2	19841	1	SCU96166	U96166 Streptococc
C 15	80	13.1	10738	1	AE006565	AE006565 Streptoco
C 16	79.8	13.1	10029	1	AE014953	AE014953 Streptoco
C 17	79.8	13.1	10738	1	AE010045	AE010045 Streptoco
C 18	79.8	13.1	50511	1	AE014154	AE014154 Streptoco
C 19	79.8	13.1	311650	1	AP005144	AP005144 Streptoco
C 20	79.2	13.0	225528	1	AL591984	AL591984 Listeria
C 21	79.2	13.0	244528	6	AX641673	AX641673 Sequence
C 22	78.2	12.8	94802	2	SPNEU1913	AL449935 Streptoco
C 23	77.6	12.7	10950	1	AE007663	AE007663 Clostridi
C 24	77.4	12.7	10687	1	AE008518	AE008518 Streptoco
C 25	76.2	12.5	10899	1	AF242367	AF242367 Lactococc
C 26	75.6	12.4	21975	1	AE014236	AE014236 Streptoco
C 27	75.4	12.4	708	6	AX570190	AX570190 Sequence
C 28	75.4	12.4	10462	1	AE007458	AE007458 Streptoco
C 29	75.4	12.4	21706	6	BD003716	BD003716 Polynucle
C 30	75.4	12.4	21706	6	BD003716	Continuation (6 of
C 31	75.4	12.4	349980	6	AX571764	AX571764 Sequence
C 32	75	12.3	12278	6	AX416895	AX416895 Sequence
C 33	74.8	12.3	12451	1	AE015017	AE015017 Streptoco
C 34	74.8	12.3	17650	1	AF397166	AF397166 Streptoco
C 35	74.8	12.3	301289	1	AE017000	AE017000 Bacillus
C 36	74.6	12.2	12834	1	U39686	U39686 Mycoplasma
C 37	74.6	12.2	13370	1	U39728	U39728 Mycoplasma
C 38	74.6	12.2	80073	6	AR300198_5	Continuation (6 of
C 39	74.6	12.2	110000	6	AR300198_0	AR300198 Sequence
C 40	74.6	12.2	250050	1	AL591978	AL591978 Listeria
C 41	74.6	12.2	349980	6	AX641667	AX641667 Sequence
C 42	74.6	12.2	349980	6	AX641668	AX641668 Sequence
C 43	74.4	12.2	12214	1	AE006299	AE006299 Lactococc
C 44	74.2	12.2	291804	1	AE017039	AE017039 Bacillus
C 45	74	12.2	753	6	AX608453	AX608453 Sequence

ALIGNMENTS

RESULT 1	AE010640	11964 bp	DNA	linear	BCT 25-MAR-2002
AE010640/c	Fusobacterium nucleatum subsp. nucleatum ATCC 25586				
LOCUS	197 of the complete genome.				
DEFINITION	AE010640 AE009951				
ACCESSION	AE010640.1 GI:19714998				
VERSION	Fusobacterium nucleatum subsp. nucleatum ATCC 25586				
KEYWORDS	Fusobacterium nucleatum subsp. nucleatum ATCC 25586				
SOURCE	Bacteria: Fusobacteria; Fusobacteriales; Fusobacteriaceae;				
ORGANISM	Fusobacterium.				
REFERENCE	1 (bases 1 to 11964)				
AUTHORS	Kapatral,V., Anderson,I., Ivanova,N., Resnik,G., Los,T.,				
	Lykidis,A., Bhattacharya,A., Bartman,A., Gardner,W., Grechkin,G.,				

Zhu, L., Vasieva, O., Chu, L., Kogan, Y., Chaga, Y., Chaga, O., Goltzman, E., Bernal, A., Larsen, N., D'Souza, M., Walunas, T., Pusch, G., Haselkorn, R., Fongstein, M., Kyripides, N. and Overbeek, R.
Genome sequence and analysis of the oral bacterium *Fusobacterium nucleatum* strain ATCC 25586
J. Bacteriol. 184 (7), 2005-2018 (2002)
2186394
J. Bacteriol. 184 (7), 2005-2018 (2002)
11889109
Kapral, V., Anderson, I., Ivanova, N., Reznik, G., Los, T., Lykidis, A., Bhattacharaya, A., Bartman, A., Gardner, W., Grechkin, G., Lykidis, A., Chu, L., Kogan, Y., Chaga, O., Goltzman, E., Bernal, A., Larsen, N., D'Souza, M., Walunas, T., Pusch, G., D., Haselkorn, R., Fongstein, M., Kyripides, N. and Overbeek, R.
Direct Submission
Submitted (13-FEB-2002) Integrated Genomics, 2201 W. Campbell Park Drive, Chicago, IL 60612, USA
Location/Qualifiers
1. .11964
/organism="Fusobacterium nucleatum subsp. nucleatum ATCC 25586"
/mol_type="genomic DNA"
/strain="ATCC 25586"
/sub_species="nucleatum"
/db_xref="ATCC:25586"
/db_xref="taxon:190304"
complement(98.772)
/gene="FN1352"
complement(98.772)
/gene="FN1352"
/codon_start=1
/transl_table=11
/product="ABC transporter ATP-binding protein"
/protein_id="AAL95548.1"
/db_xref="GI:19714993"
/translation="MDNREVLLEVNKVSIGYDLHAKVNFQVRKGEWAIMGSSGS GKSTIMVIGMDKPSIEGVIDGDITKESQNSLTIRKIKGLIFQPHLIPYTA LENVVAQYHSIDEEBALQERVLGKRAKHLPSQSGEOQRCVCIARALINSE IILADEPTGNLDEVNEKIVIEILKQLKHEGSTIVVTHLEVGDAERKILYEGKII DIIDOK"
complement(776.1978)
/gene="FN1353"
complement(776.1978)
/gene="FN1353"
/codon_start=1
/transl_table=11
/product="ABC transporter permease protein"
/protein_id="AAL95549.1"
/db_xref="GI:19715000"
/translation="MTRKQMTIKLVSVSLIRKRMIVALLAVAIGATIMSGIVTIYY DIPQLGKEFFSYANFVLPSENGKITDTEFDKINEMSTQKIVGNAPRYETIIEG CQPLITGDMIEVKNSPFWIEGWSNDENNVMGKIEISKLNLOVGETFIIEG PRAGAKVASQSDSAESKSKOLNSDPYSKKLVKGIITGGAESFIFLPSILANE ILEDDTKIDSIEADSKOLDNLANKLTADENITARPVKRTQSDIVLGLQAL VLLNIVLVILTMISVSTWMAVAERKEIGLAKGAVDSIRKEFLGEGSALGFI GGLLVGGLGVPFAEVSLSVFGRAIEFQWLPAPITIVSMIITLACLYPVKXAMIE PALVHKGK"
complement(1988.3274)
/gene="FN1354"
complement(1988.3274)
/gene="FN1354"
/codon_start=1
/transl_table=11
/product="ABC transporter permease protein"
/protein_id="AAL95550.1"
/db_xref="GI:19715001"
/translation="MIMFMRVGRGLFRQSKMLMIAFTVALGVSLATAMRWMLGVG DKVNEKLTQYANTVMHKASILDLYGLSGEGVSNKFLSEGEVPKIKOIFWGFALV DPAFLBRTGVEBSVKNKIYGTWFKHLVMPTEGVEDAGIKNLTWMEIKGWLKD DDLQEGMTIGIAGKYNKIKIGDSINIKGTNETKLTVRGLSGDDDEALIVLTKT CQLDEEDKTNIEVSALTTPDNDLAKAAQDPSLTISSEYETWCYAYSSISYIQI EVLTDVAGGNQVAESEGILTNKTEMLLILCLSSPFAISALGSLNLTATYSVIERSOE IGLIKATGKTRRRIILLILTLVLTGIFGGIFGYIAGTQIIGKTVFSSYIEPAVI

Zhu, L., Vasieva, O., Chu, L., Kogan, Y., Chaga, Y., Chaga, O., Goltzman, E., Bernal, A., Larsen, N., D'Souza, M., Walunas, T., Pusch, G., Haselkorn, R., Fongstein, M., Kyripides, N. and Overbeek, R.
Genome sequence and analysis of the oral bacterium *Fusobacterium nucleatum* strain ATCC 25586
J. Bacteriol. 184 (7), 2005-2018 (2002)
2186394
J. Bacteriol. 184 (7), 2005-2018 (2002)
11889109
Kapral, V., Anderson, I., Ivanova, N., Reznik, G., Los, T., Lykidis, A., Bhattacharaya, A., Bartman, A., Gardner, W., Grechkin, G., Lykidis, A., Chu, L., Kogan, Y., Chaga, O., Goltzman, E., Bernal, A., Larsen, N., D'Souza, M., Walunas, T., Pusch, G., D., Haselkorn, R., Fongstein, M., Kyripides, N. and Overbeek, R.
Direct Submission
Submitted (13-FEB-2002) Integrated Genomics, 2201 W. Campbell Park Drive, Chicago, IL 60612, USA
Location/Qualifiers
1. .11964
/organism="Fusobacterium nucleatum subsp. nucleatum ATCC 25586"
/mol_type="genomic DNA"
/strain="ATCC 25586"
/sub_species="nucleatum"
/db_xref="ATCC:25586"
/db_xref="taxon:190304"
complement(98.772)
/gene="FN1352"
complement(98.772)
/gene="FN1352"
/codon_start=1
/transl_table=11
/product="ABC transporter ATP-binding protein"
/protein_id="AAL95548.1"
/db_xref="GI:19714993"
/translation="MDNREVLLEVNKVSIGYDLHAKVNFQVRKGEWAIMGSSGS GKSTIMVIGMDKPSIEGVIDGDITKESQNSLTIRKIKGLIFQPHLIPYTA LENVVAQYHSIDEEBALQERVLGKRAKHLPSQSGEOQRCVCIARALINSE IILADEPTGNLDEVNEKIVIEILKQLKHEGSTIVVTHLEVGDAERKILYEGKII DIIDOK"
complement(776.1978)
/gene="FN1353"
complement(776.1978)
/gene="FN1353"
/codon_start=1
/transl_table=11
/product="ABC transporter permease protein"
/protein_id="AAL95549.1"
/db_xref="GI:19715000"
/translation="MTRKQMTIKLVSVSLIRKRMIVALLAVAIGATIMSGIVTIYY DIPQLGKEFFSYANFVLPSENGKITDTEFDKINEMSTQKIVGNAPRYETIIEG CQPLITGDMIEVKNSPFWIEGWSNDENNVMGKIEISKLNLOVGETFIIEG PRAGAKVASQSDSAESKSKOLNSDPYSKKLVKGIITGGAESFIFLPSILANE ILEDDTKIDSIEADSKOLDNLANKLTADENITARPVKRTQSDIVLGLQAL VLLNIVLVILTMISVSTWMAVAERKEIGLAKGAVDSIRKEFLGEGSALGFI GGLLVGGLGVPFAEVSLSVFGRAIEFQWLPAPITIVSMIITLACLYPVKXAMIE PALVHKGK"
complement(1988.3274)
/gene="FN1354"
complement(1988.3274)
/gene="FN1354"
/codon_start=1
/transl_table=11
/product="ABC transporter permease protein"
/protein_id="AAL95550.1"
/db_xref="GI:19715001"
/translation="MIMFMRVGRGLFRQSKMLMIAFTVALGVSLATAMRWMLGVG DKVNEKLTQYANTVMHKASILDLYGLSGEGVSNKFLSEGEVPKIKOIFWGFALV DPAFLBRTGVEBSVKNKIYGTWFKHLVMPTEGVEDAGIKNLTWMEIKGWLKD DDLQEGMTIGIAGKYNKIKIGDSINIKGTNETKLTVRGLSGDDDEALIVLTKT CQLDEEDKTNIEVSALTTPDNDLAKAAQDPSLTISSEYETWCYAYSSISYIQI EVLTDVAGGNQVAESEGILTNKTEMLLILCLSSPFAISALGSLNLTATYSVIERSOE IGLIKATGKTRRRIILLILTLVLTGIFGGIFGYIAGTQIIGKTVFSSYIEPAVI

VVPIDIALVFAVTIIGSIPAIRYLLTLKTEVLHGR"
complement(3271.4167)
/gene="FN1355"
complement(3271.4167)
/gene="FN1355"
/codon_start=1
/transl_table=11
/product="Integral membrane protein"
/protein_id="AAL95551.1"
/db_xref="GI:19715002"
/translation="WTKBEVAFGEOSFGTQSLRVGGFLGLLTIFLIALSVOKVYFR LKSSAIKALLIYLIASDFDLRGVSALARLRLFKSSNLVSNLVNVPEDKSTAYIVI LTIACISLILLFKDSRKVIGTFKNNALLRLKARLKNKHWLSLAFSLSLVSFI TVHSHITPVALTPPQIQEENMIVIPITDVEGHLRFSTIATGNNVRFIVVKK PGGSGGLGDACDTCGAGVYFERNDEIVCKRCDVVMNKSTIGFGGCGNPVFEYEIK NKKIYDKATLEKEKDRFPVGD"
4564.4722
/gene="FN1356"
4564.4722
/gene="FN1356"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAL95552.1"
/db_xref="GI:19715003"
/translation="NYDKYIKVLSIQPPFIVFYVLYNLFILWTFVHLKSPCYFF REITATLP"
4831.6006
/gene="FN1357"
4831.6006
/gene="FN1357"
/codon_start=1
/transl_table=11
/product="Transposase"
/protein_id="AAL95553.1"
/db_xref="GI:19715004"
/translation="WFLIGIDIAKLHVASCIDSTNEIVSNFKFKNDFKGFSTLLD KIKTDKNLIGLESTHYGENLINFQHFVALLINPLOTSHLRKANIRDNNDN LQSLNIAKSLFAKLNIFSEKNICFLKTRFRSLIKORSKAKIOLTLLOLLPFL ELQFLKSKHSKAIYSLLKYPSPEEIAALDDDEISLLYASKSGHFKKSKTELKS LAKTVIGKDSISLHVILIELYLDKQIKDITVKTADTVDKLTGLLSVPGISII ACAILGTNNFNFSSTKLLAFAGLPKIQSGNFVASSCRMSKSKSPILRALIF TAWNIVRHSKFNKYICLRSGKSHYNSLGHVAHVIVFTLIKINIVTQENLE"
6403.7104
/gene="FN1358"
6403.7104
/gene="FN1358"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAL95554.1"
/db_xref="GI:19715005"
/translation="MYEDNPRLENEVLIEIMKVNFLSYNSDNTTETDPENINEN VTTDDNKIYLPAPTDLEELAKGPSNMDDITLNFVVEIILENENIEGLVINPFG DLYILSKWELSELKDMKDKRLKAVEVRLKSKILISEPKHLPIMNMTIKNCCDNLG NIKKAWLEMTEDKSNWLLILDPEGDKNYIFSKISQARNYLGNWLDMLPYEDDFA RNSVQNHKAFYTKK"
7279.8766
/gene="FN1359"
7279.8766
/gene="FN1359"
/codon_start=1
/transl_table=11
/product="Dipeptide-binding protein"
/protein_id="AAL95555.1"
/db_xref="GI:19715006"
/translation="MKRLFFKVLVSLITLTFIACQKEENKEESIRTVSDIDSL NPQVVSASQILINVEGLIMPQDGVTPALASVEISEDGRTTFSIRKGVKPH NGNDMDIKDVEFSNLVSGKGNPTALPENIEKIEILDSDSHIAIYLSKPSDFIY MKEAIVPDENKHLKDIAIGTPYKLABYOKESOKLISKQNEPYMGEKAKISTVILIS PNSEINFUKLSREINFITINDPRIPELCKYILSSPSNLCILSLNPKKPPDDIE VRKAINLAIDKKNVIQLAMNGKPTIYTNMSPVMSFLWNAPEEQALPEKAKLDIBEK KLLPMEFTLKVNSKSKFYLDTAQIRQLKDIGITVNLMEIETWLSDVYTNKKYVA

```
SLAGSKMPEADAILRYVSTYPKNETFNENRYDVLVEAEKRTSNEAKOVQVYKEAQ
KILSEBOAAFLMDPNLIIATEKGIEGFQYPIPLNFALYFKK"
8776..9693
/gene="FN1360"
8776..9693
/gene="FN1360"
/codon_start=1
/transl_table=11
/product="Dipeptide transport system permease protein
dppb"
/protein_id="AAL95556.1"
/db_xref="GI:19715007"
/translation="MYIKKIFRMLLSIFSGTFSPLLELIPDPTTILGIEASAK
DLENRLQGLNLSGFSRYWNWLCGVQGLGIFSKYKEHVFNLIERLPLTISIAFI
SIFVFTMSPLSPFKTKRIKIGESILSIFSPFGLIIFMFLYGLIKWA
STGVNNTQSLILPCLVLSIPKIGWISMHLYSNLYKELREDYIKLYSNGMKVILNF
YILKNAFLPIPLTGMILLISLITGVVIEIQIPSPIGRLLVQSVLMRDPIQLGLIF
YTSFVLLNFIDILYSLDLPRIQVGEQ"
9693..10460
/gene="FN1361"
9693..10460
/gene="FN1361"
/codon_start=1
/transl_table=11
/product="Dipeptide transport system permease protein
dppc"
/protein_id="AAL95557.1"
/db_xref="GI:19715008"
/translation="MKVLYIIILVGLIIFCIAFYQYKISNFYLLKPSFOHILGT
DNLRGDFSLLLGTFSYIFIAPSALLAGILGIAJAGYEGVIDEPLRISIF
MSIPVILITIGIIVLNGEHSIILALFVLMPTRLNYRGLVKKRKHNYIKIAY
GVNHFLLRIHAPNIIPLVNPSTNFAGAILTEASGLYGLFGIOPPYPTLGNMNE
SQSYFLAPFTIPLGLMILFLVYKINQISKYQEK"
10460..11161
/gene="FN1362"
10460..11161
/gene="FN1362"
/codon_start=1
/transl_table=11
/product="Dipeptide transport ATP-binding protein dppd"
/protein_id="AAL95558.1"
Query Match 15.8%; Score 96.4; DB 1; Length 11964;
Best Local Similarity 48.7%; Pred. No. 2.9e-18;
Matches 299; Conservative 0; Mismatches 306; Indels 9; Gaps 1;
Qy 5 TAGAAATCAATGACCTCAAGAAATCTTTTGGCGTTCGGATCTTATGGCAAGGTCTCAGTC 64
Db 750 TGGAAAGTAAATAATGTGCTAAATATATATGAGATTTACATGCTTTAAAGAGTTAATT 691
Qy 65 ATAAGTTTTCACAGACACANTGACAGCACTGACTGGAGCGTCCGGTTCAGGAAATCGA 124
Db 690 TTCAAGTAAGAAAGGTAATGGGTGCAATAATGGGTTTCATCTGGTTCGGGAAATCAA 631
Qy 125 CTTTGCTCAACTGCTTGGCACACTTGACAAACCAAGTTCGGACAGATCCTTGTGAGG 184
Db 630 CTATAAGTAATATATAGTTGTATGNTAAACCAAGTATTGGTGAAGTTATTTAGATG 571
Qy 185 ATGTAGACCTTCTGAAACTCTCTAGCGGTAGAACCGGTTATACAGGAAATAACGGTGG 244
Db 570 GACAGGATATTACAAAGAAAGTCAAAATCTTTTAAACAAATAAGAGAAATAATG 511
Qy 245 GCTATTTATTTCAAGATTATCCCTTTGATTCGGACAGCAGAGTAAATTCACCTTCAGC 304
Db 510 GATTAATATTTCAACAATTCATTAAATCTTTTAACTGCTCTTGAANAATGTAATGG 451
Qy 305 TTGGCGTGGAAAAACACAAATGGGCTGAAATTCCTCAAGTACTTTCATGC-----TG 355
Db 450 TTGCTCAGTATTATCATAGTATACAGATGAAGAAGACATTACAGGCACCTTGAAGAG 391
Qy 356 TTGCTCTGAGTCGTCGAGGAAAGCCAGTTTGTGAAGTCTCTGCTGGTGGGACACAAAC 415
Db 390 TTGGACTTTAAGATAGACAAACATTTTACCTAGTCAATTATCTGCTGGTGGGACAAACA 331
```

```
Qy 416 GAACTGCGTTGCCCGGTACTGCTCAAAATCCCGGATATTTGCTGATGAACAA 475
Db 330 GAGTATGTAATAGCAAGAGCTTTAAATCAATAGCCCTGAAATAATATCTTGAGATGACAA 271
Qy 476 CCGGAGCTCTAGATTAAACAAACAGTACGTAGTTCATGAAGCATTGAGACACTCGCGG 535
Db 270 CAGGAACACCTTGATGAAGTAATGAAAAATTTGTTATAGAAATATTTAAACCACTTCATA 211
Qy 536 ACAAGAGGCCACCGTGTGTTGTCTAGCGACTCGCCCTCTTCCGAGATCAGCGGATA 595
Db 210 AAGAAGGTTCAACAATTTGTTGTAACACATGATTTAGAAGTTGGAGATGGGAGAGA 151
Qy 596 CCATTATCAAACTA 609
Db 150 GAAATAATATTA 137
RESULT 2
BX248355/c
LOCUS
DEFINITION
Corynebacterium diphtheriae gravis NCTC13129, complete genome;
segment 2/8.
ACCESSION
BX248355 BX248353
VERSION
BX248355.1 GI:38199254
KEYWORDS
complete genome.
SOURCE
Corynebacterium diphtheriae
ORGANISM
Bacteria; Actinobacteridae; Actinobacteriaceae; Corynebacterium.
REFERENCE
1 (bases 1 to 341553)
Cerdeno-Tarraga,A.M., Estratiou,A., Dover,L.G., Holden,M.T.G.,
Fallen,M., Bentley,S.D., Besra,G.S., Churcher,C., James,K.D., De
Zoya,A., Chillingworth,T., Cronin,A., Dowd,L., Feltwell,T.,
Hamlin,N., Holroyd,S., Jagels,K., Moule,S., Quail,M.A.,
Rabinowitsch,E., Rutherford,K., Thomson,N.R., Unwin,L.,
Whitehead,S. and Barrell,B.G. Parkhill,J.
The complete genome sequence and analysis of Corynebacterium
diphtheriae NCTC13129
Nucleic Acids Res. 31 (22), 6516-6523 (2003)
14602910
REFERENCE
2 (bases 1 to 341553)
Cerdeno-Tarraga,A.M.
Direct Submission
Submitted (03-OCT-2003) Cerdeno-Tarraga A.M., submitted on behalf
of the Pathogen Sequencing Unit, Sanger Institute, Wellcome Trust
Genome Campus, Hinxton, Cambridge CB10 1SA E-mail:
amct@sanger.ac.uk
FEATURES
Location/Qualifiers
1..341553
/organism="Corynebacterium diphtheriae"
/mol_type="genomic DNA"
/strain="NCTC13129"
/db_xref="taxon:1717"
/note="biotype gravis"
/complement(351..1163)
/locus_tag="DIP0382"
/complement(351..1163)
/locus_tag="DIP0382"
/note="Similar to Streptomyces coelicolor hypothetical
29.8 kDa protein 28CD46.17C TR:Q9FCG6 (EMBL:AL391406) (271
aa) fasta scores: E(): 3e-26, 47.92% id in 265 aa"
/codon_start=1
/transl_table=11
/product="Conserved hypothetical protein"
/protein_id="CAE4886.1"
/db_xref="GI:38199255"
/translation="MADAHNLRAITFGHGREIPGVITRGTTGFRNLRCDRWAHPLI
QQALRTSDPIALDVGYCAFTTVEWARLRTWEDDVRTGLEIDPERVLPPRDGVT
PLGGFELAGYOPTLVRAFNVLRQYDVQVMDAMWMTSRLOPGGFFVEGTCDELGR
STWVLLNHAGPSLTLCDPDPVERPSDIAERLPKILIHNVAGSKIHDVLLRTDEAW
DECAFWASGPRVWRARALMDMGIPMSIIRRPVRDNLTPVPSIVDPTT"
complement(1179..11640)
gene
CDS
```

```

/locus_tag="DIP0383"
complement(1179..1640)
/locus_tag="DIP0383"
/note="Low similarity to Mycobacterium tuberculosis
hypothetical 19.1 kDa protein Rv0481c or M70499 or
MTCV20G3.07c SW:Y481 MYCTU (Q11147) (174 aa) fasta scores:
E(): 0.028, 23.92% id in 163 aa"
/codon_start=1
/transl_table=11
/product="Conserved hypothetical protein"
/protein_id="CAE48887.1"
/db_xref="GI:38199256"
/translacion="MEKVHAALTNPAYWDFVTKNLSPEPGVHEFTTEANGAVATLPE
VLPBLLPEATRAIMISQALVKRVVTVVGLNGTSAPLSYADVKGTVDFKGETSMNG
LGNTTLLSYANELIIVNIFLMGAATPEKVGQALADLPQNEGDLTSKWSKNL"
1676..2809
/locus_tag="DIP0384"
1676..2809
/locus_tag="DIP0384"
/note="Similar to Mycobacterium tuberculosis
UDP-N-acetylglucosamine reductase MurB or
Rv0482 or M70500 or MTCV20G9.08 SW:MURB MYCTU (Q11148)
(369 aa) fasta scores: E(): 1.7e-39, 51.27% id in 353 aa,
and to Escherichia coli
UDP-N-acetylglucosamine reductase MurB or
B3972 SW:MURB_ECOLI (P08373) (342 aa) fasta scores: E():
3.1e-14, 31.69% id in 325 aa"
/codon_start=1
/transl_table=11
/product="Putative peptidoglycan synthesis related
protein"
/protein_id="CAE48888.1"
/db_xref="GI:38199257"
/translacion="MSCIQTTFNEVLEKSLNHLIDGIRPQIEADDAFVAVTFPAEILT
LHGGTTPAAVRCRSQSVVVRLLDAHQIPLLIIVGGNSLIVADGEIPLVAVILD
CDLISVTDTRGVAAEAGVWDVRLVCVDAAGLIGLCSGIPGSAGATPVQNVAYG
AEISDLVSVLLPRTAGTANVEVPAADLELAVRYSLNLFKTRGVVLGITQLHWDGWS
APLRGEARVLGHEGPHPAVQVRAVLGLRAGKGMYNADHDHTWAGSFFNPIVP
ESVGDHRSVVGDSMPCFAAGEGVKKLSAALIDRAGFGHGGPGGRVSLSTPKHFL
ALTRNGNATTDLDALAEVRGVNDAPFGLLEPEPVWVGVS1"
1787..2317
/locus_tag="DIP0384"
/note="HMPfam hit to PF02215,
UDP-N-acetylglucosamine reductase, FAD-binding
domain"
2387..2797
/locus_tag="DIP0384"
/note="HMPfam hit to PF02873,
UDP-N-acetylglucosamine reductase, C-terminal
domain"
complement(2806..3171)
/locus_tag="DIP0385"
complement(2806..3171)
/locus_tag="DIP0385"
/note="Similar to Bacillus subtilis Yjqa protein TR:Q34593
(EMBL:D29110) (125 aa) fasta scores: E(): 5e-11, 35.51% id
in 107 aa"
/codon_start=1
/transl_table=11
/product="Conserved hypothetical protein"
/protein_id="CAE48889.1"
/db_xref="GI:38199258"
/translacion="METASITETWTFVQVPIPEDVANLVPGEDAVVAFKTRDRAIF
TKRLIVRDSQGLRKQKVEVYSLPYSSVNMWSTENAGTFDPAEVLWTRAGKIKKI
GRDLVRSLMLAQCIIG"
complement(2809..2820)
/locus_tag="DIP0385"
/note="ScanRegExp hit to PS00294, Prenyl group binding
site (CAAX box)"
complement(3174..4880)
/locus_tag="DIP0386"
complement(3174..4880)
/locus_tag="DIP0386"

```

```

/note="Similar to Escherichia coli
long-chain-fatty-acid--CoA ligase FadD or OldD or B1805
SW:LCFA_ECOLI (P29212) (561 aa) fasta scores: E():
7.5e-62, 34.46% id in 560 aa. Possible duplication of the
downstream CDS: Similar to DIP0387 (566 aa) fasta scores:
E(): 1.2e-164, 73.488% identity in 562 aa overlap"
/codon_start=1
/transl_table=11
/product="Putative ligase"
/protein_id="CAE48890.1"
/db_xref="GI:38199259"
/translacion="MGAFDDKAWLOHYGWTWPASLDYGDITLVDIYNNLAVNADKPA
TYFFQRTQYAELEDAQRAAAAGAFGRPGDRVAIVLPNCPOHIAAFYAIQLIGAT
VVEHNPITYAHELEBGLFQDHGARVAIAWDKAATLEKRTDTSLETIISVNMTEAMPR
IQOLALRIPPIPIAAKRLQAFADNTVPWSTLIGNAICGNGKRTPEPKVKDIALI
LVTSGTGTGKALLSHGNLSVNCMPGKAVFGLGDQPERFLAALPMFAYGEMTAVGT
LGVFVGAEMVLLPAQIPLIMKMKNTPTWLPGVPTLYEKIVABATAGVIEIKGIRN
SFGSASTLPVTVEDWENITGCLLVGEGYGLTETSPVLCGNMNGNRBPYIGVPPDT
EIRIANPNLDIETMEDGAGEVLRGPQIFQYLKPKETAATFHGDWFRIGDMQVME
EDGFRILVSRIKEIITGGFNVPYPAEVEALIAHPDVAAGVGRPKDGEDVAVCI
VLKDAALDPGLKTHCKRLTRYKVPRTFYHFEALNKDQLGKVRREVQQTLLERLG
EK"
complement(3405..4724)
/locus_tag="DIP0386"
/note="HMPfam hit to PF00501, AMP-binding enzyme"
complement(4182..4208)
/locus_tag="DIP0386"
/note="PrintScan hit to PR00154, AMP-binding signature"
complement(4194..4229)
/locus_tag="DIP0386"
/note="ScanRegExp hit to PS00455, Putative AMP-binding
domain signature"
complement(4209..4244)
/locus_tag="DIP0386"
/note="PrintScan hit to PR00154, AMP-binding signature"
complement(4942..6642)
/locus_tag="DIP0387"
complement(4942..6642)
/locus_tag="DIP0387"
/note="Similar to Escherichia coli
long-chain-fatty-acid--CoA ligase FadD or OldD or B1805
SW:LCFA_ECOLI (P29212) (561 aa) fasta scores: E():
7.8e-63, 34.69% id in 565 aa. Possible duplication of the
upstream CDS: Similar to DIP0386 (568 aa) fasta scores:
E(): 1.2e-164, 73.488% identity in 562 aa overlap"
/codon_start=1
/transl_table=11
/product="Putative ligase"
/protein_id="CAE48891.1"
/db_xref="GI:38199260"
/translacion="MSAVETKKLHYDPWTPHSLDYGDITLVDIYNNLANKDKPA
TYFFQRTSYGLDQKQVTAAGRALGVRSQDRVAIALPNCPOHIAVAVLMLGAT
VVEHNPITYAHELESFQDQHGARVAIVDKAATLEKRNTPLETIVSVNMWTAMPA
VQOALRUPIMIRKRRDALTAASANTPDIWELVGSALGNGDDIVPCPEVTKKSIAL
ILYTSGTGTGKAGLSHGNLFANIMQGWKVGIGDQDERLLAALPMFAYGIVTIVL
NLAFTVGGELVLLPAQIPLIMKMKHTPTWPGVPTLYEKIVABERDGISGVR
NSFGASISPLVTVARWALTGCLLVGEGYGLTETSPVLCGNMNGNRBPYIGVPPD
TEIRIANPNLDIETMPDQGEVLRGPQVFGVSLNNPEATAKSFHGQWRTGVDGIM
EDGFRILVARIKEVITGGFNVPYPAEVEEVLNHPDITDATVGLPSGDGSESVAA
ITLGAALDPGLKAYCENLTRYKVPRTFYHFEELPKDQLGKVRREVQQTALASKK
"
complement(5164..6486)
/locus_tag="DIP0387"
/note="HMPfam hit to PF00501, AMP-binding enzyme"
complement(5953..5988)
/locus_tag="DIP0387"
/note="ScanRegExp hit to PS00455, Putative AMP-binding
domain signature"
6726..8009
/locus_tag="DIP0388"
6726..8009
/locus_tag="DIP0388"
/note="Similar to Streptomyces coelicolor putative

```

glycosyl transferase GLGA TR:Q9X9U5 (EMBL:AJ243803) (387

Query Match 15.1%; Score 91.8; DB 1; Length 341553;
Best Local Similarity 49.2%; Pred. No. 1.9e-16;
Matches 280; Conservative 0; Mismatches 277; Indels 12; Gaps 1;

QY 53 AAGTCTCAGTCATAGTTTTCACAGAAACAATGACAGCACTGACTGGACGTCGCGTT 112
Db 69796 AGGGAATGATCTCACTATTTTACCTGAGAACTGTTGCAATTTGCGAAATCTGTT 69737

QY 113 CAGGAATACGATTTGCTCAACTGCTCTTGACACATGACAAACCAAGTCCGACAGA 172
Db 69736 CAGGAATACATCTACTGACTGCTTAAAGTATCTCGCACCGCATGAAGATCTA 69677

QY 173 TCCTTGTGAGAGTATAGACCTTCTGAAACTCTCTAGCGTAAGCAACCGTTATACAGGA 232
Db 69676 TCTACCTAGCCAATCAGAACATCAGCCCACTCGCCACGAAAGTCTGCTGAGATCAGGA 69617

QY 233 AATATACGCTGGCTATTTATTTCAAGATATGCTTGAATCCCGACAGACAGTTAAAT 292
Db 69616 GAAATCAGCAAGTTTCATTTTCAAGACTTTAAACCTCATTAGTCTCCCTGACGTCGCCG 69557

QY 293 TCAACCTTCT-----AGCTTGGGTGGGAAACACAAATGCGCTGAAATCTCTC 340
Db 69556 ACAACATCGGCTTCTTCAAGCTTGTGCGCAACCCCAACCAACACGAATCGACT 69497

QY 341 AAGTACTTCACTGTTGCTTGTGCTGAGTCTGAGTCTGAGGAAAGCCAGTTTGAATCTCTG 400
Db 69496 CAGTCTTGTGAAAGTGGGCTGTGAGAGTGTGCAATAATATCCAAACCAATATCCG 69437

QY 401 GTGCGCAACAAACGAACATGCTGTTGGCGGTGAGTCTGCTCAAAATCCCGAATATTC 460
Db 69436 GAGACACACACACAGCGGTAGTATCGAGTCTCTGTGAGTCTCTCCACAGATCTCT 69377

QY 461 TGGTGTATGACCAACCGGAGCTCTAGATTAAACACAGTGAAGTCTAGTATGAGAGCAT 520
Db 69376 TTGCGGACGAACCGACCGGAGCTAGACGTTCACACCGCAATATCTTAAACTAC 69317

QY 521 TGAGAGACTCGCGCAAGAGCGCCACCGTGTGTTGTTGTCAGCACTGCCCTCTTCC 580
Db 69316 TTCAGAAACGGTCACCAAGAGAACTTCTCTAGTTATGTCACACATGATCTGGATTTG 69257

QY 581 GAGATCAGCGGATACCATTTATCAAACTA 609
Db 69256 CAGCATCAGTGCACCGGTGTTGTGTGCTA 69228

RESULT 3

BX571657/c
LOCUS BX571657 347961 bp DNA linear BCT 08-SEP-2003
DEFINITION Wolinella succinogenes, complete genome; segment 1/7.
ACCESSION BX571657 BX571656
VERSION BX571657.1 GI:34482172
KEYWORDS complete genome.
SOURCE Wolinella succinogenes
ORGANISM Wolinella succinogenes
Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
Helicobacteraceae; Wolinella.
REFERENCE 1
AUTHORS Schuster, S.C., Baar, C., Eppinger, M., Raddatz, G., Simon, J.M.,
Lanz, C., Klumek, O., Nandakumar, R., Gross, R., Rosinus, A.,
Keller, H., Jagtap, P., Linke, B., Meyer, F. and Lederer, H.
Complete genome sequence and analysis of Wolinella succinogenes
Unpublished
REFERENCE 2
AUTHORS Schuster, S.C.
TITLE Direct Submission
JOURNAL Submitted (15-MAY-2003) Max-Planck Institut for Developmental
Biology, Spemannstr. 35, 72076 Tuebingen, GERMANY
LOCATION/Qualifiers
1. .347961
/organism="Wolinella succinogenes"
/mol_type="genomic DNA"

/strain="DSMZ 1740"
/db_xref="taxon:844"
1. .1314
/gene="dnaA"
/locus_tag="WS0000"
1. .1314
/gene="dnaA"
/locus_tag="WS0000"
/codon_start=1
/transl_table=11
/product="dnaA, probable chromosomal replication initiator
protein"
/protein_id="CAE09174.1"
/db_xref="GI:34482173"
/translation="MLGPTTLKOLKDEISTLEYEYIKQLRYDEASRTDLAVYIAPN
QIANWKTGYGERLAHFAELSTGPRKIEIKGLKKOVKSSPAAGSKGQKSTIL
NPSFTDVSQVNSRFAEVSQNAKKQIAYNPLLYGGTGLKTHLLNSIGNYNV
AGKGVIVYTSQFLNDYLYHRRNTMDFRDKYRACDYLIDDDYQFGGPKQIEEF
PHTFELHNKMKQIVLTDKTPKQIAGLEERLKSFEWGWSDIQPPELETINIKK
KCEFGIVLSNEIISVIATNDNIREIEGIIKLNAYANLMNOBITLQFAKNVLEQ
IKERENITLENII EWVSKELNVKPAEIKSKGRSKNIANARIVIFLARTLTTPNMP
LAQFGMKDHSVSAMKTVKTEIEDANFKMVEELKNKIKSRN"
1474. .2541
/gene="DNAN"
/locus_tag="WS0001"
1474. .2541
/gene="DNAN"
/locus_tag="WS0001"
/EC_number="2.7.7.7"
/codon_start=1
/transl_table=11
/product="DNA POLYMERASE III, BETA CHAIN"
/protein_id="CAE09175.1"
/db_xref="GI:34482174"
/translation="MNVIPKSSLENIITNLOPLEKKDAQSITSHVYFKAGSELTP
QATDYEIGLLAKTSVOIKEGIAITVNGKKLIDIVRRKKEGIEILQSEVILHKKQEK
SFKLPMNAEFSPFSIEHPKIEINSLHVLQSMKALIPKQINPNKPELNGALLD
IKESFNVAIDTRRLAVIKYENPSINKLIVPKKAIKVEIKQIFFEVEHYSETHL
ILKSNLYFTFKLISGKPYDERIIPKTHNTMLPKKIEIKLAINSLNDKIVTF
TKBELFESLSESEAKTQVETGTITKELTGLINSKYLLDLQIDTSEFKIGINE
SNTPIKSENFSTVINEIVM"
2561. .4879
/gene="gyrB"
/locus_tag="WS0002"
2561. .4879
/gene="gyrB"
/locus_tag="WS0002"
/EC_number="5.99.1.3"
/codon_start=1
/transl_table=11
/product="DNA GYRASE SUBUNIT B"
/protein_id="CAE09176.1"
/db_xref="GI:34482175"
/translation="MNOQYSADNIKVLKGLAEVRKPGMYIGTNNVNGLHLYEVV
DNSIDEAMAGHNTISITLTKEGSAIIDNNGRGIPOVMHTPENLPAATVVLTVLHAG
KFDKDSYKVGGLHGVGVVVALSKLVMTRKNGVYTCPSGIPKTKGLDITGS
KNGCTTIEFPDDEIFBVTVDREVLRARFKELAYLNRNITTHFKDERDNFESYHFE
GGLQRFVHLNKKVLVSQVFSFSDNESDTEVEIALAINDGDEKLVSVNNIKTEGG
THEAGFRAGLSRAITSIDANANREKNDKITGDVREGLVAVIVSVKMFQFEQTK
KLGCSFVKPIQKLYEKIAKFEENPNDAKAIMQALMAARGAARAKARELTRK
DSLMTGTPKGLADQCKDPSISELYVEGDSAGSAGQGRDVFQAILPLKGLVAV
EKSRLDKILKSDIKMTALCGGIGEEFDINIRYOKIITMDADVDSGHIOPLLIT
FFRYLPKPIEAGVLYIAQPLVFFKKGKELVLRDEKALSEVLIENGIEAFEGVG
TPDLLEILRASHVRLTLEKRPFAMIDVVKVMIEHRDLISLMSLYIKLEFVQK
KGNILNKRVEEKIHLIVQTHEGLVDLIDDSLFTPLPFEAYYVNNILRLDSFL
GKEDILILEIEESAKKGATIQRYKGLGEMNPELQWETTTMPENRRLKVKQITDDME
ASDVSLFMGDEVEPRRAYIQDHAKQVHKLDV"
4890. .5264
/locus_tag="WS0003"
4890. .5264
/locus_tag="WS0003"
/note="HYPOTHETICAL PROTEIN CUI724C"
/codon_start=1


```

ACCESSION      AP003515
VERSION        AP003515.1  GI:15076712
KEYWORDS
SOURCE
ORGANISM       Clostridium perfringens str. 13
               Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
               Clostridium.
REFERENCE
AUTHORS        Shimizu,T., Ohtani,K., Hirakawa,H., Ohshima,H., Yamashita,A.,
               Shiba,T., Ogasawara,N., Hattori,M., Kuhara,S. and Hayaashi,H.
TITLE          Complete genome sequence of Clostridium perfringens, an anaerobic
               flesh-eater
JOURNAL        Proc. Natl. Acad. Sci. U.S.A. 99 (2), 996-1001 (2002)
MEDLINE        21664373
PUBMED        11792842
REFERENCE      2 (bases 1 to 54310)
AUTHORS        Shimizu,T.
TITLE          Direct Submission
JOURNAL        Submitted (12-APR-2001) Tohru Shimizu, Institute of Basic Medical
               Sciences, University of Tsukuba, Department of Microbiology; 1-1-1
               Tennoudai, Tsukuba, Ibaraki 305-8575, Japan
               (E-mail:tshimizu@md.tsukuba.ac.jp, Tel:81-298-53-3354,
               Fax:81-298-53-3354)
FEATURES
source         1. 54310
               /organism="Clostridium perfringens str. 13"
               /mol_type="genomic DNA"
               /strain="13"
               /db_xref="taxon:195102"
               /plasmid="pCP13"
               /note="anaerobic pathogen for gas gangrene"
gene           940..1692
               /gene="soj"
               /note="250 aa, similar to p1r:I40444 Spo0A activation
               inhibitor soj from Bacillus subtilis (253 aa); 37%
               identity in 250 aa overlap
CDS            PCP01
               /codon_start=1
               /transl_table=11
               /product="Soj protein"
               /protein_id="BAB62438.1"
               /db_xref="GI:15076713"
               /translation="MKKISVENIKGVAKTTSANFGACILEEKGDVLLVDLPQSNL
               TKLFKXSMEDVSIADVLLDKNLDIEKVIKKTDFENIDILPSNVTIAPAEKILLDVN
               RSQNRKAKALEEDKIDYCLIDCPALNMTIVNLCASDEVIVPIKIDKFDLGDLE
               YLLDSIEIKDFENPNLNFKGFITMDSSTTVNKKVTKQELKSVLGEKMFNTSIHQNIK
               VVESTFECPVFSSKKARASLNYKDLKSEIP"
               /length=1751
               /gene="parB"
               /note="426 aa, similar to gp:AF300944_3 presumptive ParB
               protein from Lactococcus lactis subsplactis (242 aa); 30%
               identity in 266 aa overlap
               PCP02"
               /codon_start=1
               /transl_table=11
               /product="parB protein"
               /protein_id="BAB62439.1"
               /db_xref="GI:15076714"
               /translation="MAKKFSISEGLMNGISKNTRKVEEFOAKENFKIEVINIDIKRN
               EKNFYVIDISELAEIDKLNGLNMLVVRKLDNMYELISGERRYALSGLVNEGKNE
               KIRDIANDLNLSTQVGRYERLNKLIPELKAVIEQGNLTIANASEFSLSENGRV
               ILSLIDTWNMQEAVDLNKKLKHHEKELELKKAYEKELELKEELKEEKKNQVSK
               LKSENEKLLKLSNNLEERKEIEGGIKIEFEKKNKVEIKNFVQLNKLIDNSFNKLSQI
               TKAEKNLQKQKDLKSLKESKNNEVDIKNTKENFVLVQNLKIDNSFNKLSQI
               NMKKNVKAETTKAEFELEKYQKEISDLLKKL"
               /length=3147
               /gene="PCP03"
               /note="120 aa, no significant homology"
               /codon_start=1
               /transl_table=11
               /product="hypothetical protein"
               /protein_id="BAB62440.1"
               /db_xref="GI:15076715"
               /translation="NEKILAEKRINISFYKRNKALVTTLLPDKWLEVIGITENERE
               KPFYEDAKIKISKEQSEAKEKTIISFKSTKTYLNKKWLEVLGVSDEDSKCIEL
               RKFDITLVKDNGRDILDI"
               /length=3773
               /gene="PCP04"
               /note="83 aa, similar to p1r:T14710 probable transposase
               from Yersinia pestis (402 aa); 44% identity in 50 aa
               overlap
               truncated"
               /codon_start=1
               /transl_table=11
               /product="probable transposase"
               /protein_id="BAB62442.1"
               /db_xref="GI:15076717"
               /translation="MSKDFLNYIKEQNFNDSNNVLSLKDLFNDVLQEVVSAELDDM
               LGEKMLNQPVNSNGYSKTIKSELGTITILNLYRN"
               /length=4040
               /gene="PCP05"
               /note="60 aa, similar to gp:AF143819_1 transposase-like
               protein from Escherichia coli (402 aa); 38% identity in 60
               aa overlap
               truncated"
               /codon_start=1
               /transl_table=11
               /product="probable transposase"
               /protein_id="BAB62443.1"
               /db_xref="GI:15076718"
               /translation="MISKHQNNINIEKLNLYASGITTRDVAGIKALYDIETSAE
               TVSNITNRIPELVSEW"
               /length=4588
               /gene="PCP06"
               /note="52 aa, similar to p1r:T43600 probable transposase
               from Yersinia pestis (105 aa); 56% identity in 50 aa
               overlap
               truncated"
               /codon_start=1
               /transl_table=11
               /product="probable transposase"
               /protein_id="BAB62444.1"
               /db_xref="GI:15076719"
               /translation="WLYTTNVIESLNSQFRKFTKTLIPFNDVSLIKMLYLATEKVNK
               KWTRNYPN"
               /complement(5169..5804)
               /gene="PCP07"
               /note="211 aa, similar to gp:AP001508_4 ABC transporter
               (ATP-binding protein) from Bacillus halodurans (213 aa);
               49% identity in 214 aa overlap"
               /codon_start=1
               /transl_table=11
               /product="probable ABC transporter"
               /protein_id="BAB62445.1"
               /db_xref="GI:15076720"
               /translation="WNIEISLNKKYFDKVFKDFSLIKKGEMIALSGSGGCKST
               LINMGLTEKFDSEIIDDGVKINKSLANKFLREKISYLFQNFALVDEETVEENL
               RLAKHTIKTKTBEERIRCLKFLVGLGCKNTYIELSGEQKQVALARMLKPFSEI
               ILADEPTGSLDENRDIISLLKELNESGKTIIVTHDNYVAKQADRIPTL"

```

gene complement (5804. .7966)
/gene="PCP08"
CDS complement (5804. .7966)
/gene="PCP08"
/note="720 aa, similar to gp:AP001508.3 BH0280 Gene
product from Bacillus halodurans (713 aa); 23% identity in
661 aa overlap"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB62446.1"
/db_xref="GI:15076721"
/translation="MKKVALIILFIITILSPRGVSVRNHTFEMKLNKLNLENF
EVTISIDRENYNTAYESI KSLDEYNGNIPFSEVDINEDVRKVINGVESKNETNH
LPIVSGRFFHKNDISLTSIDETCCQGVNDPFGKNIHEIRTIKSKLDINTEN
LFIIVENQBLDILDLSESIIIVQRVWGSSQYNTETKILILVVCFIGLIFMIF
YQVLGSKIGIQILGHSTFVMLKERLLEVLIVLAVTVLLVFFNFKNLSLF
WKFMLETCIYIMIFIVSVIPIYIVKWTETKQYILPELGNDESIIQSFSEMEK
ASILIFSNALDLSIGIGYKXVWSEKQYILPELGNDESIIQSFSEMEK
BRAYLVENKOGADLADPNRYEPTMEAKOMLPEYVRETIIVPNYLKHKVYDVE
GNINISEDEKDRILLPEKRYNPEKEILEYGYNSOEPGSSSTCSHKTAGKLNVE
QKXIIWKSKQKIFSLLOVNPENGVTDPIVSLTESNDKLVSVYKILIGYNSPF
KIRANSEEVINGLEKYDMSVILIDPYNLYDNVASTIINIKAKRVIIIFAIIVILAV
ISIIILQNTSLYFNQNKIIVKHLGRLIYRYMYFIIMVLITWTCPFLAIALSUTKD
INIIVFTLIVLIVELFIIFINSLKKNLIKVIKGY"
complement (8043. .8354)
/gene="PCP09"
CDS complement (8043. .8354)
/gene="PCP09"
/note="103 aa, no significant homology"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB62447.1"
/db_xref="GI:15076722"
/translation="MKRNKLIKTIASLLVLSGLTALAYTHSSDKFEASLPGLFS
QAQSKFVCGQKRRATRVKVTTLVEAKDIKAKLTAHAQTKYIKGVTEWNSYAH
L"
gene 8779. .9012
CDS /gene="PCP10"
8779. .9012
/gene="PCP10"
/note="77 aa, similar to probable transposase from
Yersinia pestis plasmid pMT1 (402 aa); 25% identity in 158
aa overlap"
/codon_start=1
/transl_table=11
/product="probable transposase"
/protein_id="BAB62448.1"
/db_xref="GI:15076723"
Query Match 13.8%; Score 83.8; DB 1; Length 54310;
Best Local Similarity 47.7%; Pred. No. 3.7e-14;
Matches 296; Conservative 0; Mismatches 307; Indels 18; Gaps 1;
QY 1 ATGATAGAAATCAATGACTCAGAAATCTTTGGCGTTCCGATCTATGCAAGTCTC 60
DB 5798 ATTATAGAAATTCAAATTTAAATAAAAGTATTTTGATAGGTAATTTTAAAGATTTT 5739
QY 61 AGTCATAGTTTATACCGAGAACATGACGACCTGACTGGAGCGTCCGGTTACAGGAAA 120
DB 5798 TCTTTGAGTATATAAAAGGTGAGATGATGCTATTAGTGAGAGAGTGTTGTGGAAG 5679
QY 121 TCGATTTTCTCAATGCTCTGGACACTTGACAAACCAAGTTCGGACAGATCCTTGTC 180
DB 5678 AGTACCCCTTTTAAATATGATAGGCTTGATAGAAAATTTGATTCAGGAAATAATTATT 5619
QY 181 GAGGATGATAGCTTCTGAACTCTCTACGGGTAGCAACGGTTATACAGGAAAAATACG 240
DB 5618 GACGTGTATAAAATATAAATAATAGCAATTAGCAATAGTTTITTAGGGAAG 5559
QY 241 GTGGGCTATTTATTCAGATTTATGCTTGATTCCTCCGACGAGACAGTAAATTCACCTT 300

Db 5558 ATAAGTTACTTATTTCAAAATTTTCCTTGGTAGATGAAGAACTGTAGAGGAGATTTA 5499
QY 301 CAGCTTGGCGTGA--AAACCAAAATGCCCTGAAATTCCTCAA 342
DB 5498 AGACTTGCATTAAGCACACAAATTAATAACAAAGATTTGAAGAAGAAATATAAGA 5439
QY 343 GTACTTTCATGCTTGGTTCGAGTCGTCGAGGAGAAAGCCAGTTTGTGAACCTCTCGT 402
DB 5438 TGTTTAAATTTGTTGGATTAGGAGATGTCMAAAATACTATATATAGATTAGTGTG 5379
QY 403 GCGCAACCAACCAAGAACTGCGTTGGCCCGGCTCTCAAAAATCCCGCAATATTCG 462
DB 5378 GGTGAGCAACAGAGAGTAGCTATCGTAGATTAAATGTTAAACCAAGTGAATATTTTA 5319
QY 463 GCTGATGAACCAACCGAGCTCTAGATTACAAACAGTGCAGCTAGTATAGAACATTG 522
DB 5318 GCAGATGAACCAACAGGCTCATTAGTAGAAGAAATAGAGATTTATATATCTTTATTA 5259
QY 523 AGAGCACTCGCCGACAAAGGCGCCACCGTTGTTGTGTACGCACTCGCCCTCTTCCA 582
DB 5258 AAGAACTTAATGAAGTGGAAACAAATAATATCGTACTCATGATAATTATGTTGCA 5199
QY 583 GAATCAGCGATACCATTAATC 603
DB 5198 AAACAGCAGATAGGATAATC 5178
RESULT 5
AE015943/c 301439 bp DNA linear BCT 06-FEB-2003
LOCUS Clostridium tetani E88, section 8 of 10 of the complete genome.
DEFINITION AE015943 AE015927
ACCESSION AE015943.1 GI:28204047
VERSION
KEYWORDS Clostridium tetani E88
SOURCE Clostridium tetani E88
ORGANISM Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
Clostridium.
REFERENCE 1 (bases 1 to 301439)
AUTHORS Bruggemann, H., Baumer, S., Fricke, W.F., Wierzer, A., Liesegang, H.,
Decker, I., Herzberg, C., Martinez-Arias, R., Merkl, R., Henne, A. and
Gottschalk, G.
TITLE The genome sequence of Clostridium tetani, the causative agent of
tetanus disease
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (3), 1316-1321 (2003)
MEDLINE 22457253
PUBMED 12552129
REFERENCE 2 (bases 1 to 301439)
AUTHORS Bruggemann, H., Baumer, S., Fricke, W.F., Wierzer, A., Liesegang, H.,
Decker, I., Herzberg, C., Martinez-Arias, R., Merkl, R., Henne, A. and
Gottschalk, G.
TITLE Direct Submission
JOURNAL Submitted (31-OCT-2002) Goettingen Genomics Laboratory, Institute
of Microbiology and Genetics, Georg-August University,
Grisebachstr. 8, Goettingen 37077, Germany
FEATURES
source
1. .301439
/organism="Clostridium tetani E88"
/mol_type="genomic DNA"
/strain="Massachusetts"
/sub_strain="E88"
/db_xref="taxon:212717"
92. .778
/locus_tag="CTC01981"
92. .778
/locus_tag="CTC01981"
/codon_start=1
/transl_table=11
/product="L-serine dehydratase beta subunit"
/protein_id="AA036487.1"
/db_xref="GI:28204048"
/translation="MSRNFSVFDIVGPIWIGPSSSHTAGAAFLAKLAASIAIGEKIKKY


```
NEIHPMILAMNEIGYDWTWLTGNIHFNFYGVPTLEKVASQFSGTMLCGNVYKQDERLKG
PYNVEKAGVKVGLIGTWTNTRWDSNENLKYRVTDPEETKVKVSEIKQKVDVMA
VWNSGEYGNKNSAVSLANACPELVAHAEHKAJDGATYNNVNLVKNKLAQ
IAKIDILIKDGGYIVSDKIDVTQSLIIMWOGKTKKNYESDKLEKLSYHKVA
LDDANQIIGELKGNLVPEDVVGIPQSOETPMNLINVEQMYITKANVSAAAFR
SNAMKCKIKKSDASLIYFQDNTLYLLEVTQKLDYMSWSSYNTYKPGDLTVSF
NEDVGVNDFSGVKEIIDIISKEPGNMTINLMDNSLVKDTETELKLIIVNRRASH
LLNPSGIFKDGSLPKLIEKDALDGPIDRLIGRYKEVKKGVITPQMNKNKISGN
WDSAKREAVKLINLEGKIKVQSDARTPNVASITEKOLENAKNKAVKPDVKKPTTK
PDSKSDKDKVIRPADKKYVTVYKSGDCLYLQKQNVSYKTKAKANNIKWNLIIVG
ORLIIPAK"

Query Match      13.7%; Score 83.6; DB 1; Length 301439;
Best Local Similarity 48.3%; Pred. No. 6.6e-14;
Matches 282; Conservative 0; Mismatches 284; Indels 18; Gaps 1;

QY 1 ATGATAGAAATCAATGACCTCAAGAAATCTTTGGCGTTCGGATCTTATGCGAAGGTCTC 60
DB 125348 ATTGTAAGATGGAGATATTAACATAAAATTTTGGAGATAAATATCTAAATAATTT 125289
QY 61 AGTCATAGTTTATACAGGAACATGACACACTGACTGGAGCGTCCGGTTCAGGAAA 120
DB 125288 TCAGTAGATATACAGATGGAGAACTCTAGCTGTAAGCAAGTGGAAAGTGGAAA 125229
QY 121 TCGACTTTGCTCAACTGTTGGACACTTGACAAACCAAGTTCGGACAGATCCTTGTC 180
DB 125228 TCTACTATATTAATATATAGGCGTTTGTAGAGGTTTGTGATCTGGAAGCTTATTTA 125169
QY 181 GAGGATGATAGCTTCTGAAACTCTCTAGCGTGAAGCAACGGTTATACAGAAAAATACG 240
DB 125168 GATGATGATGAATATATAAATAAATTCATCGAATCAATAGATATTAAGAGAAAA 125109
QY 241 GTGGGCTATTTATTTCAAGATTATGCTTGTATTCCTGACAGACAGATTAATTCACCTT 300
DB 125108 ATAGGATATTTATTTCAAAACTTTGCAATAGTAGATGAAGAAACAGTATATATAATCTA 125049
QY 301 CAGCTTGGGCTGG-----AAAAACAACAATGGCCTGAAATTCCTCAA 342
DB 125048 CATTGAGCACTTAATATATGTTAAGAAAAATAAAGAAGAAAGATGATGATGATAAAAT 124989
QY 343 GTACTTTCATGCTGTGTGGTTCGAGTCGTTGAGGAAAAAGCCAGTTTGTGAATCTCTGGT 402
DB 124988 GTATTAAGCAAAATGAATTTAGAGGATGATGAAGAGAAAAATCTTTGAGTTAAGTGGC 124929
QY 403 GCGGAACACACAGCACTCGTGTGCGCGGCTACTGCTCAAAATCCCGAATATATCTG 462
DB 124928 GGAGAGCAACAAAGAGTTTCAATAGCAAGATTTATTATTAACCAAGCAAGATATTTA 124869
QY 463 GCTGATGAACCAACCGAGCTCTAGATTAAACAAACAGTGTAGTCTAGTATAGAACATG 522
DB 124868 GCTGATGAACCAACCGGATCATTAGATGCTAAAAATAGAGACTTGTATTATCTATTTA 124809
QY 523 AGAGCACTCGCGCAAAAGCCGACCGTTGTGTGTCTACGCA 566
DB 124808 AATAAACTCAACAAAGAGGAAAAACAGTAATAGTAGTAACCTCA 124765

RESULT 6
AE000977 Archaeoglobus fulgidus DSM 4304 linear BCT 17-MAR-2003
LOCUS Archaeoglobus fulgidus DSM 4304 section 130 of 172 of the complete
DEFINITION genome.
ACCESSION AE000977 AE000782
VERSION AE000977.1 GI:2689300
KEYWORDS Archaeoglobus fulgidus DSM 4304
SOURCE Archaeoglobus fulgidus DSM 4304
ORGANISM Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
Archaeoglobaceae; Archaeoglobus.
REFERENCE 1 (bases 1 to 10575)
AUTHORS Klenk, H.P., Clayton, R.A., Tomb, J., White, O., Nelson, K.E.,
Ketchum, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, J.D.,
Richardson, D.L., Kierlavage, A.R., Graham, M.D., Adams, M.D., Loftus, B.,
Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S.,
Kirkness, E.F., Dougherty, B.A., McKenney, K., Badger, J.H., Glodek, A.,
Peterson, S., Reich, C.I., McNeil, L.K., Badger, J.H., Glodek, A.,
Zhou, L., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L.,
Utterback, T., Cotton, P.W., Spriggs, T., Artach, P., Kaine, B.P.,
Sykes, S.M., Sadow, P.W., D'Andrea, K.P., Bowman, C., Fujii, C.,
Garland, S.A., Mason, T.M., Olsen, G.J., Fraser, C.M., Smith, H.O.,
Woese, C.R. and Venter, J.C.
The complete genome sequence of the hyperthermophilic,
sulphate-reducing archaeon Archaeoglobus fulgidus
Nature 390 (6658), 364-370 (1997)
9389475
2 (bases 1 to 10575)
Klenk, H.P., Clayton, R.A., Tomb, J.-P., White, O., Nelson, K.E.,
Ketchum, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, J.D.,
Richardson, D.L., Kierlavage, A.R., Graham, M.D., Adams, M.D., Loftus, B.,
Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S.,
Kirkness, E.F., Dougherty, B.A., McKenney, K., Badger, J.H., Glodek, A.,
Peterson, S., Reich, C.I., McNeil, L.K., Badger, J.H., Glodek, A.,
Zhou, L., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L.,
Utterback, T., Cotton, P.W., Spriggs, T., Artach, P., Kaine, B.P.,
Sykes, S.M., Sadow, P.W., D'Andrea, K.P., Bowman, C., Fujii, C.,
Garland, S.A., Mason, T.M., Olsen, G.J., Fraser, C.M., Smith, H.O.,
Woese, C.R. and Venter, J.C.
Direct Submission
Submitted (15-DEC-1997) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
In order to show the genes in ascending order on the genome, the
origin of this version has been moved by TIGR to position 2093570
of the original version and the opposite strand is shown from the
original version.
On Dec 16, 1997 this sequence version replaced gi:2648725.
Location/Qualifiers
1. .10575
/organism="Archaeoglobus fulgidus DSM 4304"
/mol_type="genomic DNA"
/db_xref="taxon:224325"
/complement(318..1667)
/gene="AF1813"
/complement(318..1667)
/gene="AF1813"
/feature="similar to GP-2225877 percent identity: 47.43;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="TBP-interacting protein TIP49"
/protein_id="AAB89434.1"
/db_xref="GI:2648730"
/translation="MAGEIRBITOTFERISAHSHIRGLDENLRAKDVADGLVGOKR
AREAGVIRLLIKSGRMAGRILMAGPOTGKTATNAVALSKELGKIDIPVOVSASEFY
SAEMKKEALIQAMRAIGRIRRETIVLEGEVGLDYNWNPYNPTCKIPESATLT
LATKDEKRTSVGGRLAMQFTQGIIEVGDVIVIDKETGKIGKLGKAKKIDGDD
EVVVPDGVKVEKEFTYVVTLDLDEANARSTISFSPSPRIDNEVRBADEQV
KRVPEGRALVPGVLFIDETHLMDIELFAFMNAMESMAPITIIASNRGFSRIGT
DIVAPGIPLDLRLITTPYREBKTIETIRAAESGIMLSNEAKELTDICEK
TSLRYAVQLLAPAYEFAKNSGKVELEDERAASIFADVSSQSSAYLKKWEEKMLGM"
1751..2107
/gene="AF1814"
1751..2107
/gene="AF1814"
/notes="hypothetical protein; identified by GeneMark;
putative"
/codon_start=1
/transl_table=11
/product="A. fulgidus predicted coding region AF1814"
/protein_id="AAB89439.1"
/db_xref="GI:2648735"
/translation="MKLRGHLIICLOFFREGYSGFVENVKIMERLESGETVEVVA
GVDDICVAPYNRGYCGKDETSEERVREMDKAAVLLGVSEKTTTREVAREIERKVLNE
WRVFCDCQWKEVCYG"
2144..3418
/gene="AF1815"
```

2144. .3418
 /gene="Af1815"
 /note="similar to GB:L77117 SP:Q58131 PID:1591438 percent identity: 36.19; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="acetylornithine aminotransferase (argD-2)"
 /protein_id="AAB89433.1"
 /db_xref="GI:2648729"
 /translation="MGFTCRGCKAEIIIEAFKSHVSPYKAKFFSMVGIDFVPARREGVW
 WYLDGKRLMDCHNGGVFNGLHRHPIVKTIVLALDELDTGNHLLISEQRAIAEKL
 AELMPGDISRTVFGVGGEAIDFAIKLARGHTGRKKIYAKGGYHGTGFGALAAEKL
 IYKFFELAPGVPEVFGDAEAVEKAVDSDYAAVLFTETPATGLMPLPDPFRRVRE
 RKPFCGLMDEVDTGRTGKMGWISHYKVPDVTIVTAKGLGGVYPSIFSCFKEG
 LDDFWAENFIIVSTFGAEICVGVAAEVLBITRESGFLENVRKTGEALSEILGKLKD
 EYDFVIRQKGLFIQIKMVEEGWGPLLSISCYHSGILAVYANNDSWQFLPPLIVG
 KEVDYVIREKLAGEAIEASQRREWVEFKRL"
 3415. .4317
 /gene="Af1816"
 3415. .4317
 /gene="Af1816"
 /note="hypothetical protein; identified by GeneMark; putative"
 /codon_start=1
 /transl_table=11
 /product="A. fulgidus predicted coding region AF1816"
 /protein_id="AAB89438.1"
 /db_xref="GI:2648734"
 /translation="MKVDVEVLNFEERLNPAPKENIKIYGGEISTVFELPELPGIA
 LKRIINFKSVQEVRYIEVYFEYIELLEGAGLRVPDSDATFERRDGTVLVIAQKMW
 GNVGKVIQTAKDAEAEVFGKVLAEKWKVDENAEEREGLEIGIDGQISWALMEE
 IYVDTSTPMVKNGIHOLDTEFLQCPGIRILLKKIFLOQILDRIYDFDKRVVDM
 IANLYKESREDLIPIAINLANEFPFLQCPGIRILLKKIFLOQILDRIYDFDKRVVDM
 VHRFLITKALFGRYEYILPGKRR"
 4457. .4966
 /gene="Af1817"
 4457. .4966
 /gene="Af1817"
 /note="similar to PID:868019 GB:U00096 SP:Q59431
 PID:1743672 PD:1742880 percent identity: 24.55;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="transcriptional regulatory protein, TetR family"
 /protein_id="AAB89432.1"
 /db_xref="GI:2648728"
 /translation="MTDOSVRIIEAURLYMKPPHVEISIEEAKVSKSLIFVYFH
 ESKQKLLSEAVNHAFKQMEFNPSVEEVDYIGFIAERREFIEFMMYALSQVRIE
 ELERMFGLAEKVASLFEGRCHPRETAIALMAMLDGLSIYSLYFLGKLEKYRIAME
 FVESRRVA"
 4963. .6399
 /gene="Af1818"
 4963. .6399
 /gene="Af1818"
 /note="hypothetical protein; identified by GeneMark; putative"
 /codon_start=1
 /transl_table=11
 /product="A. fulgidus predicted coding region AF1818"
 /protein_id="AAB89437.1"
 /db_xref="GI:2648733"
 /translation="MRFVLALLAFALAAQATQPVITLNHHTIPENFSIGYFIISD
 ISNAGENARGSLVTINEEGFVLINPDSEKTSVYLDGLDSAGTSAGIKKAYAEK
 SGLYKALASLVNRESGEFVIGVAYDLPFIYAGSTTIEIPDSRSEAYMLRN
 LGGDVESAVIILNSDFTTSOTALIGNSRGCTKKVVLGLKADSDAVAGVYAKIIVDY
 RSPEKGVASVPLVIVPEVTVLEVSFPEKVVSDFSLSKVAITAKSKSLGSLT
 ASLEAREPDGKKYKVTGAEKAVINFSKLPRESGEYDKVLVIRTESFAKQFSPV
 LYISDYGISIDLAGVTVFQRVVEGDSFKLSIQVENSEGDAKAVANLLIPGLEG
 DRSGVIGLSGDSATATFMRKAKGSKQKIVAEITYMDKTFEKHKAEREFEVYVFA
 6400. .7095
 /gene="Af1819"
 /note="similar to GB:L77117 SP:Q58131 PID:1591438 percent identity: 36.19; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="acetylornithine aminotransferase (argD-2)"
 /protein_id="AAB89433.1"
 /db_xref="GI:2648729"
 /translation="MGFTCRGCKAEIIIEAFKSHVSPYKAKFFSMVGIDFVPARREGVW
 WYLDGKRLMDCHNGGVFNGLHRHPIVKTIVLALDELDTGNHLLISEQRAIAEKL
 AELMPGDISRTVFGVGGEAIDFAIKLARGHTGRKKIYAKGGYHGTGFGALAAEKL
 IYKFFELAPGVPEVFGDAEAVEKAVDSDYAAVLFTETPATGLMPLPDPFRRVRE
 RKPFCGLMDEVDTGRTGKMGWISHYKVPDVTIVTAKGLGGVYPSIFSCFKEG
 LDDFWAENFIIVSTFGAEICVGVAAEVLBITRESGFLENVRKTGEALSEILGKLKD
 EYDFVIRQKGLFIQIKMVEEGWGPLLSISCYHSGILAVYANNDSWQFLPPLIVG
 KEVDYVIREKLAGEAIEASQRREWVEFKRL"

Qy 472 CCACCGGAGCTCTAGATTATCAAAACAGTGTAGCTAGTATCATAGAACGATTCAGACGACTC 531
|||||
Db 6886 CCAACTGGCAGTCTGGACTTCGAACTGGAAAGGCTGTCTGAAGTCTATGAGAGAGATA 6945
|||||

Qy 532 GCGACAAAGCGCCACCGTGTGTT 558
|||||

Db 6946 ACAGGAAGGAGGAATAACCTTTGT 6972
|||||

RESULT 7
LOCUS AE007598 10415 bp DNA linear BCT 27-JUL-2001
DEFINITION Clostridium acetobutylicum ATCC824 section 86 of 356 of the complete genome.
VERSION AE007598 AE001437
KEYWORDS AE007598.1 GI:15023711
SOURCE
ORGANISM Clostridium acetobutylicum
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae; Clostridium.
REFERENCE 1 (bases 1 to 10415)
Nolling, J., Breton, G., Omelchenko, M. V., Markarova, K. S., Zeng, Q., Gibson, R., Lee, H. M., Dubois, J., Qiu, D., Hitti, J., Wolf, Y. I., Tatusov, R. L., Sabathe, F., Doucette-Stamm, L., Soucaille, P., Daly, M. J., Bennett, G. N., Koonin, E. V. and Smith, D. R.
Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum
J. Bacteriol. 183 (16), 4823-4838 (2001)
21359325
11466286
REFERENCE 2 (bases 1 to 10415)
Childress, D., Zeng, Q. and Smith, D. R.
Direct Submission
Submitted (24-JUL-2001) GTC Sequencing Center Production, Finishing, and Bioinformatics teams, Genome Therapeutics Corp., 100 Beaver Street, Waltham, MA 02453-8443, USA
Location/Qualifiers
1. 10415
/organism="Clostridium acetobutylicum"
/mol_type="genomic DNA"
/strain="ATCC 824"
/db_xref="ATCC:824"
/db_xref="taxon:1488"
122..724
/gene="CAC0821"
122..724
/gene="CAC0821"
/codon_start=1
/transl_table=11
/product="Transcriptional regulator (Tetr/AcrR family)"
/protein_id="AAK78797.1"
/db_xref="GI:15023712"
/translation="WQYLEKREKRIIDEALKEFLRGLTYNASTRTTVQAGTSVGNF KYFGKEDLYEKIDIPVYKLIYHOFSEVNEQAQIFELMKSVIEIPEBND ELSMNSGTSKNCRCQIFIDFITNTVTSFEVLYMNRKTIKDNFIELAHNV DQAIILRKQKNTQVRKLIIVNLIDIFYGNIIDKIDIOEI"
744..1448
/gene="CAC0822"
744..1448
/gene="CAC0822"
/codon_start=1
/transl_table=11
/product="ABC transporter, ATP-binding protein"
/protein_id="AAK78798.1"
/db_xref="GI:15023713"
/translation="WESPIELRDVREKIYMGVEVTKAVDGVSFSDIGEFVIVLGASG AKSTNLNLGMDQASGSIISVQNEI.SNVMKMLTKYREDIDIGFVQFVNLQVNL ALENVLAEIKENPMDPVKILSVGLKDRMNPFSQSGEQORVSIARALAKPKL LICDETGALDINIGKALKLADISKQYNMTVILITHNSAIAFIADKILHVKNRVE SVKKNQTPVPIESIEW"
1450..4800
/gene="CAC0823"

CDS

1450..4800
/gene="CAC0823"
/codon_start=1
/transl_table=11
/product="Predicted membrane protein"
/protein_id="AAK78799.1"
/db_xref="GI:15023714"
/translation="MKKTFKFLPKSITSLRFLSILIIIAVGSFYVGVGRASSPDM KMSADKYFNDNFMDKLIISLTGLTDKDVKRLSKVKAAGSKSDMDAVIENGKDTI VLVBSLPKQGVNVKQVVRGRPLSKNTLAEVERFLKNGKFGIDVLKLSGNDTNI EDRKLCSPKIVGTVTSPLVISAQOISSVGNVNGKGVVIMPEAFKNDVYTSIYVKI NEPSKDSIAQNAKYKASIVNVKELKTFVKEGCEARRKQIVKSDSKIDKSDKLSA SKKEAKDFIDAHKKLDDAVYKNDGKVLQKQNDKFTQITSNREIIONKSEIQOS EYILNSKLADAKSKSTLEGAEQIVLQQLNDQKHAAEGISAALGRKAVSDAKNKL KNPDSITCKEIIYNNLNDVYTDIQGEFDDMYESLSSNSLEKISSYFDIKGAKAF DSAQSEISCKQTIDNEKALISGEVQLQNALQINAGKDKIAKETELSKNKGLE KLNAARKIDDSKEINEKTEETNANNKFNNAEGRKRNKRLNKEVPEPTVY LGRNENLGVNYSDSDRIDNIGKVFPLIFLVAALVSLTMTVMVOENETGCTPKA LGSPKVIIVSHYLVISFTASIIIGSLGIIIGFEVPEPLIVNAGSLVAVPHYSFDSV TSIEAVITAVFTTASAIAYLDELKEVPASLKPPEKPKKILLERMTFIWRFPSF TKKTARNIRYKQRFMTVIGIACTGLMTGEGIKSGVIGAVEKQFSSVYXVQGT TINKIDDSERKEIKKNSDSNKISLVLFYSKNATVRKEDNNEDINIVPEKKQL NEYELFNDSELKLSDDGVIVTEKIAKLMNKKGVDFNITVDNRILIRVIFDITKHY IQHYIYMSPKYKVTGENAFQNAFGLLKSSSSAENTTSKISGINKIAGVFPKN TRDFNKSTLSNVLVAVSAGLAVVIVNLININERRELATIKVLFQYDKE LAMVYIRENIIITIGSLVIGVGMLLDVFVITTTETNMVFLRTVEPLVILISFLT IAFSVINIAMVRRFYKIDMIESLKSAAE"
4906..5226
/gene="CAC0824"
4906..5226
/gene="CAC0824"
/codon_start=1
/transl_table=11
/product="Thioredoxin"
/protein_id="AAK78800.1"
/db_xref="GI:15023715"
/translation="WESLQSIKEIRKYIENNELVLMYFRSDNSGICVYALDPLEAISK KYKSISSVKNVTGNLPSVPADFKITVPEVLIIFMNGSIVIEHFKYINFRIEIEVLRY YKCI"
complement(5283..6416)
/gene="CAC0825"
complement(5283..6416)
/gene="CAC0825"
/codon_start=1
/transl_table=11
/product="Endoglucanase family 5"
/protein_id="AAK78801.1"
/db_xref="GI:15023716"
/translation="MFRFLFSKVKILFLALTLGLTYLTLNLTVAATNNTTKTKKIT SLQIVKQKVLGLDGLNSDFPNETDMGNPKTKSMITKIQAGFNTVPIPTWNSHI GPADYITDNLDRVQEVNVAVDNHYVIINLHEDSWLIPTIANKDAATAEIKTL WTQISNPKYNDLSLIPETNPELIGTPEEWYGGTSEARDVNVNLAANAIRLTG AKNSRRIPITYAASDKISINGLKIPIPNNDKILISLHAYIPYEMVNGVSSWGS AKDKSLDNLAIYKFKGKGVVIGCEPASVKNKNENORASLAKYVVSAAKRSIA PIWMDNGVSYSGAKDSYGFNRNLTWRWECPIVHALTGLKVKK"
complement(6748..7860)
/gene="CAC0826"
complement(6748..7860)
/gene="CAC0826"
/codon_start=1
/transl_table=11
/product="Endoglucanase family 5"
/protein_id="AAK78802.1"
/db_xref="GI:15023717"
/translation="MFKNLFSKAKISLLTFAFLGTSLSGAYSACAATVPHDMSQOI VNDMKVQNLGNTLDSAPDEGTGNPKTKTKMIDIKIKEAGFNTVPIPTWNSHIGAGP SYTIDQALNRQEVNVAVDNHYVAILNTHDTSWIIPYIANKDAATAEIKTL WTQISNPKYNDLSLIPETNPELIGTPEEWYGGTSEARDVNVNLAANAIRLTG AKNSRRIPITYAASDKISINGLKIPIPNNDKILISLHAYIPYEMVNGVSSWGS AKDKSLDNLAIYKFKGKGVVIGCEPASVKNKNENORASLAKYVVSAAKRSIA PIWMDNGVSYSGAKDSYGFNRNLTWRWECPIVHALTGLKVKK"
complement(6748..7860)
/gene="CAC0826"
complement(6748..7860)
/gene="CAC0826"
/codon_start=1
/transl_table=11
/product="Endoglucanase family 5"
/protein_id="AAK78802.1"
/db_xref="GI:15023717"
/translation="MFKNLFSKAKISLLTFAFLGTSLSGAYSACAATVPHDMSQOI VNDMKVQNLGNTLDSAPDEGTGNPKTKTKMIDIKIKEAGFNTVPIPTWNSHIGAGP SYTIDQALNRQEVNVAVDNHYVAILNTHDTSWIIPYIANKDAATAEIKTL WTQISNPKYNDLSLIPETNPELIGTPEEWYGGTSEARDVNVNLAANAIRLTG AKNSRRIPITYAASDKISINGLKIPIPNNDKILISLHAYIPYEMVNGVSSWGS AKDKSLDNLAIYKFKGKGVVIGCEPASVKNKNENORASLAKYVVSAAKRSIA PIWMDNGVSYSGAKDSYGFNRNLTWRWECPIVHALTGLKVKK"
complement(6748..7860)
/gene="CAC0826"
complement(6748..7860)
/gene="CAC0826"
/codon_start=1
/transl_table=11
/product="Endoglucanase family 5"
/protein_id="AAK78802.1"
/db_xref="GI:15023717"
/translation="MFKNLFSKAKISLLTFAFLGTSLSGAYSACAATVPHDMSQOI VNDMKVQNLGNTLDSAPDEGTGNPKTKTKMIDIKIKEAGFNTVPIPTWNSHIGAGP SYTIDQALNRQEVNVAVDNHYVAILNTHDTSWIIPYIANKDAATAEIKTL WTQISNPKYNDLSLIPETNPELIGTPEEWYGGTSEARDVNVNLAANAIRLTG AKNSRRIPITYAASDKISINGLKIPIPNNDKILISLHAYIPYEMVNGVSSWGS AKDKSLDNLAIYKFKGKGVVIGCEPASVKNKNENORASLAKYVVSAAKRSIA PIWMDNGVSYSGAKDSYGFNRNLTWRWECPIVHALTGLKVKK"
complement(6748..7860)
/gene="CAC0826"
complement(6748..7860)
/gene="CAC0826"
/codon_start=1
/transl_table=11
/product="Endoglucanase family 5"
/protein_id="AAK78802.1"
/db_xref="GI:15023717"
/translation="MFKNLFSKAKISLLTFAFLGTSLSGAYSACAATVPHDMSQOI VNDMKVQNLGNTLDSAPDEGTGNPKTKTKMIDIKIKEAGFNTVPIPTWNSHIGAGP SYTIDQALNRQEVNVAVDNHYVAILNTHDTSWIIPYIANKDAATAEIKTL WTQISNPKYNDLSLIPETNPELIGTPEEWYGGTSEARDVNVNLAANAIRLTG AKNSRRIPITYAASDKISINGLKIPIPNNDKILISLHAYIPYEMVNGVSSWGS AKDKSLDNLAIYKFKGKGVVIGCEPASVKNKNENORASLAKYVVSAAKRSIA PIWMDNGVSYSGAKDSYGFNRNLTWRWECPIVHALTGLKVKK"
complement(6748..7860)
/gene="CAC0826"
complement(6748..7860)
/gene="CAC0826"
/codon_start=1
/transl_table=11
/product="Endoglucanase family 5"
/protein_id="AAK78802.1"
/db_xref="GI:15023717"
/translation="MFKNLFSKAKISLLTFAFLGTSLSGAYSACAATVPHDMSQOI VNDMKVQNLGNTLDSAPDEGTGNPKTKTKMIDIKIKEAGFNTVPIPTWNSHIGAGP SYTIDQALNRQEVNVAVDNHYVAILNTHDTSWIIPYIANKDAATAEIKTL WTQISNPKYNDLSLIPETNPELIGTPEEWYGGTSEARDVNVNLAANAIRLTG AKNSRRIPITYAASDKISINGLKIPIPNNDKILISLHAYIPYEMVNGVSSWGS AKDKSLDNLAIYKFKGKGVVIGCEPASVKNKNENORASLAKYVVSAAKRSIA PIWMDNGVSYSGAKDSYGFNRNLTWRWECPIVHALTGLKVKK"
complement(6748..7860)
/gene="CAC0826"
complement(6748..7860)
/gene="CAC0826"
/codon_start=1
/transl_table=11
/product="Endoglucanase family 5"
/protein_id="AAK78802.1"
/db_xref="GI:15023717"
/translation="MFKNLFSKAKISLLTFAFLGTSLSGAYSACAATVPHDMSQOI VNDMKVQNLGNTLDSAPDEGTGNPKTKTKMIDIKIKEAGFNTVPIPTWNSHIGAGP SYTIDQALNRQEVNVAVDNHYVAILNTHDTSWIIPYIANKDAATAEIKTL WTQISNPKYNDLSLIPETNPELIGTPEEWYGGTSEARDVNVNLAANAIRLTG AKNSRRIPITYAASDKISINGLKIPIPNNDKILISLHAYIPYEMVNGVSSWGS AKDKSLDNLAIYKFKGKGVVIGCEPASVKNKNENORASLAKYVVSAAKRSIA PIWMDNGVSYSGAKDSYGFNRNLTWRWECPIVHALTGLKVKK"
complement(6748..7860)
/gene="CAC0826"
complement(6748..7860)
/gene="CAC0826"
/codon_start=1
/transl_table=11
/product="Endoglucanase family 5"
/protein_id="AAK78802.1"
/db_xref="GI:15023717"
/translation="MFKNLFSKAKISLLTFAFLGTSLSGAYSACAATVPHDMSQOI VNDMKVQNLGNTLDSAPDEGTGNPKTKTKMIDIKIKEAGFNTVPIPTWNSHIGAGP SYTIDQALNRQEVNVAVDNHYVAILNTHDTSWIIPYIANKDAATAEIKTL WTQISNPKYNDLSLIPETNPELIGTPEEWYGGTSEARDVNVNLAANAIRLTG AKNSRRIPITYAASDKISINGLKIPIPNNDKILISLHAYIPYEMVNGVSSWGS AKDKSLDNLAIYKFKGKGVVIGCEPASVKNKNENORASLAKYVVSAAKRSIA PIWMDNGVSYSGAKDSYGFNRNLTWRWECPIVHALTGLKVKK"
complement(6748..7860)
/gene="CAC0826"
complement(6748..7860)
/gene="CAC0826"
/codon_start=1
/transl_table=11
/product="Endoglucanase family 5"
/protein_id="AAK78802.1"
/db_xref="GI:15023717"
/translation="MFKNLFSKAKISLLTFAFLGTSLSGAYSACAATVPHDMSQOI VNDMKVQNLGNTLDSAPDEGTGNPKTKTKMIDIKIKEAGFNTVPIPTWNSHIGAGP SYTIDQALNRQEVNVAVDNHYVAILNTHDTSWIIPYIANKDAATAEIKTL WTQISNPKYNDLSLIPETNPELIGTPEEWYGGTSEARDVNVNLAANAIRLTG AKNSRRIPITYAASDKISINGLKIPIPNNDKILISLHAYIPYEMVNGVSSWGS AKDKSLDNLAIYKFKGKGVVIGCEPASVKNKNENORASLAKYVVSAAKRSIA PIWMDNGVSYSGAKDSYGFNRNLTWRWECPIVHALTGLKVKK"
complement(6748..7860)
/gene="CAC0826"
complement(6748..7860)
/gene="CAC0826"
/codon_start=1
/transl_table=11
/product="Endoglucanase family 5"
/protein_id="AAK78802.1"
/db_xref="GI:15023717"
/translation="MFKNLFSKAKISLLTFAFLGTSLSGAYSACAATVPHDMSQOI VNDMKVQNLGNTLDSAPDEGTGNPKTKTKMIDIKIKEAGFNTVPIPTWNSHIGAGP SYTIDQALNRQEVNVAVDNHYVAILNTHDTSWIIPYIANKDAATAEIKTL WTQISNPKYNDLSLIPETNPELIGTPEEWYGGTSEARDVNVNLAANAIRLTG AKNSRRIPITYAASDKISINGLKIPIPNNDKILISLHAYIPYEMVNGVSSWGS AKDKSLDNLAIYKFKGKGVVIGCEPASVKNKNENORASLAKYVVSAAKRSIA PIWMDNGVSYSGAKDSYGFNRNLTWRWECPIVHALTGLKVKK"
complement(6748..7860)
/gene="CAC0826"
complement(6748..7860)
/gene="CAC0826"
/codon_start=1
/transl_table=11
/product="Endoglucanase family 5"
/protein_id="AAK78802.1"
/db_xref="GI:15023717"
/translation="MFKNLFSKAKISLLTFAFLGTSLSGAYSACAATVPHDMSQOI VNDMKVQNLGNTLDSAPDEGTGNPKTKTKMIDIKIKEAGFNTVPIPTWNSHIGAGP SYTIDQALNRQEVNVAVDNHYVAILNTHDTSWIIPYIANKDAATAEIKTL WTQISNPKYNDLSLIPETNPELIGTPEEWYGGTSEARDVNVNLAANAIRLTG AKNSRRIPITYAASDKISINGLKIPIPNNDKILISLHAYIPYEMVNGVSSWGS AKDKSLDNLAIYKFKGKGVVIGCEPASVKNKNENORASLAKYVVSAAKRSIA PIWMDNGVSYSGAKDSYGFNRNLTWRWECPIVHALTGLKVKK"
complement(6748..7860)
/gene="CAC0826"
complement(6748..7860)
/gene="CAC0826"
/codon_start=1
/transl_table=11
/product="Endoglucanase family 5"
/protein_id="AAK78802.1"
/db_xref="GI:15023717"
/translation="MFKNLFSKAKISLLTFAFLGTSLSGAYSACAATVPHDMSQOI VNDMKVQNLGNTLDSAPDEGTGNPKTKTKMIDIKIKEAGFNTVPIPTWNSHIGAGP SYTIDQALNRQEVNVAVDNHYVAILNTHDTSWIIPYIANKDAATAEIKTL WTQISNPKYNDLSLIPETNPELIGTPEEWYGGTSEARDVNVNLAANAIRLTG AKNSRRIPITYAASDKISINGLKIPIPNNDKILISLHAYIPYEMVNGVSSWGS AKDKSLDNLAIYKFKGKGVVIGCEPASVKNKNENORASLAKYVVSAAKRSIA PIWMDNGVSYSGAKDSYGFNRNLTWRWECPIVHALTGLKVKK"
complement(6748..7860)
/gene="CAC0826"
complement(6748..7860)
/gene="CAC0826"
/codon_start=1
/transl_table=11
/product="Endoglucanase family 5"
/protein_id="AAK78802.1"
/db_xref="GI:15023717"
/translation="MFKNLFSKAKISLLTFAFLGTSLSGAYSACAATVPHDMSQOI VNDMKVQNLGNTLDSAPDEGTGNPKTKTKMIDIKIKEAGFNTVPIPTWNSHIGAGP SYTIDQALNRQEVNVAVDNHYVAILNTHDTSWIIPYIANKDAATAEIKTL WTQISNPKYNDLSLIPETNPELIGTPEEWYGGTSEARDVNVNLAANAIRLTG AKNSRRIPITYAASDKISINGLKIPIPNNDKILISLHAYIPYEMVNGVSSWGS AKDKSLDNLAIYKFKGKGVVIGCEPASVKNKNENORASLAKYVVSAAKRSIA PIWMDNGVSYSGAKDSYGFNRNLTWRWECPIVHALTGLKVKK"
complement(6748..7860)
/gene="CAC0826"
complement(6748..7860)
/gene="CAC0826"
/codon_start=1
/transl_table=11
/product="Endoglucanase family 5"
/protein_id="AAK78802.1"
/db_xref="GI:15023717"
/translation="MFKNLFSKAKISLLTFAFLGTSLSGAYSACAATVPHDMSQOI VNDMKVQNLGNTLDSAPDEGTGNPKTKTKMIDIKIKEAGFNTVPIPTWNSHIGAGP SYTIDQALNRQEVNVAVDNHYVAILNTHDTSWIIPYIANKDAATAEIKTL WTQISNPKYNDLSLIPETNPELIGTPEEWYGGTSEARDVNVNLAANAIRLTG AKNSRRIPITYAASDKISINGLKIPIPNNDKILISLHAYIPYEMVNGVSSWGS AKDKSLDNLAIYKFKGKGVVIGCEPASVKNKNENORASLAKYVVSAAKRSIA PIWMDNGVSYSGAKDSYGFNRNLTWRWECPIVHALTGLKVKK"
complement(6748..7860)
/gene="CAC0826"
complement(6748..7860)
/gene="CAC0826"
/codon_start=1
/transl_table=11
/product="Endoglucanase family 5"
/protein_id="AAK78802.1"
/db_xref="GI:15023717"
/translation="MFKNLFSKAKISLLTFAFLGTSLSGAYSACAATVPHDMSQOI VNDMKVQNLGNTLDSAPDEGTGNPKTKTKMIDIKIKEAGFNTVPIPTWNSHIGAGP SYTIDQALNRQEVNVAVDNHYVAILNTHDTSWIIPYIANKDAATAEIKTL WTQISNPKYNDLSLIPETNPELIGTPEEWYGGTSEARDVNVNLAANAIRLTG AKNSRRIPITYAASDKISINGLKIPIPNNDKILISLHAYIPYEMVNGVSSWGS AKDKSLDNLAIYKFKGKGVVIGCEPASVKNKNENORASLAKYVVSAAKRSIA PIWMDNGVSYSGAKDSYGFNRNLTWRWECPIVHALTGLKVKK"
complement(6748..7860)
/gene="CAC0826"
complement(6748..7860)
/gene="CAC0826"
/codon_start=1
/transl_table=11
/product="Endoglucanase family 5"
/protein_id="AAK78802.1"
/db_xref="GI:15023717"
/translation="MFKNLFSKAKISLLTFAFLGTSLSGAYSACAATVPHDMSQOI VNDMKVQNLGNTLDSAPDEGTGNPKTKTKMIDIKIKEAGFNTVPIPTWNSHIGAGP SYTIDQALNRQEVNVAVDNHYVAILNTHDTSWIIPYIANKDAATAEIKTL WTQISNPKYNDLSLIPETNPELIGTPEEWYGGTSEARDVNVNLAANAIRLTG AKNSRRIPITYAASDKISINGLKIPIPNNDKILISLHAYIPYEMVNGVSSWGS AKDKSLDNLAIYKFKGKGVVIGCEPASVKNKNENORASLAKYVVSAAKRSIA PIWMDNGVSYSGAKDSYGFNRNLTWRWECPIVHALTGLKVKK"
complement(6748..7860)
/gene="CAC0826"
complement(6748..7860)
/gene="CAC0826"
/codon_start=1
/transl_table=11
/product="Endoglucanase family 5"
/protein_id="AAK78802.1"
/db_xref="GI:15023717"
/translation="MFKNLFSKAKISLLTFAFLGTSLSGAYSACAATVPHDMSQOI VNDMKVQNLGNTLDSAPDEGTGNPKTKTKMIDIKIKEAGFNTVPIPTWNSHIGAGP SYTIDQALNRQEVNVAVDNHYVAILNTHDTSWIIPYIANKDAATAEIKTL WTQISNPKYNDLSLIPETNPELIGTPEEWYGGTSEARDVNVNLAANAIRLTG AKNSRRIPITYAASDKISINGLKIPIPNNDKILISLHAYIPYEMVNGVSSWGS AKDKSLDNLAIYKFKGKGVVIGCEPASVKNKNENORASLAKYVVSAAKRSIA PIWMDNGVSYSGAKDSYGFNRNLTWRWECPIVHALTGLKVKK"
complement(6748..7860)
/gene="CAC0826"
complement(6748..7860)
/gene="CAC0826"
/codon_start=1
/transl_table=11
/product="Endoglucanase family 5"
/protein_id="AAK78802.1"
/db_xref="GI:15023717"
/translation="MFKNLFSKAKISLLTFAFLGTSLSGAYSACAATVPHDMSQOI VNDMKVQNLGNTLDSAPDEGTGNPKTKTKMIDIKIKEAGFNTVPIPTWNSHIGAGP SYTIDQALNRQEVNVAVDNHYVAILNTHDTSWIIPYIANKDAATAEIKTL WTQISNPKYNDLSLIPETNPELIGTPEEWYGGTSEARDVNVNLAANAIRLTG AKNSRRIPITYAASDKISINGLKIPIPNNDKILISLHAYIPYEMVNGVSSWGS AKDKSLDNLAIYKFKGKGVVIGCEPASVKNKNENORASLAKYVVSAAKRSIA PIWMDNGVSYSGAKDSYGFNRNLTWRWECPIVHALTGLKVKK"
complement(6748..7860)
/gene="CAC0826"
complement(6748..7860)
/gene="CAC0826"
/codon_start=1
/transl_table=11
/product="Endoglucanase family 5"
/protein_id="AAK78802.1"
/db_xref="GI:15023717"
/translation="MFKNLFSKAKISLLTFAFLGTSLSGAYSACAATVPHDMSQOI VNDMKVQNLGNTLDSAPDEGTGNPKTKTKMIDIKIKEAGFNTVPIPTWNSHIGAGP SYTIDQALNRQEVNVAVDNHYVAILNTHDTSWIIPYIANKDAATAEIKTL WTQISNPKYNDLSLIPETNPELIGTPEEWYGGTSEARDVNVNLAANAIRLTG AKNSRRIPITYAASDKISINGLKIPIPNNDKILISLHAYIPYEMVNGVSSWGS AKDKSLDNLAIYKFKGKGVVIGCEPASVKNKNENORASLAKYVVSAAKRSIA PIWMDNGVSYSGAKDSYGFNRNLTWRWECPIVHALTGLKVKK"
complement(6748..7860)
/gene="CAC0826"
complement(6748..7860)
/gene="CAC0826"
/codon_start=1
/transl_table=11
/product="Endoglucanase family 5"
/protein_id="AAK78802.1"
/db_xref="GI:15023717"
/translation="MFKNLFSKAKISLLTFAFLGTSLSGAYSACAATVPHDMSQOI VNDMKVQNLGNTLDSAPDEGTGNPKTKTKMIDIKIKEAGFNTVPIPTWNSHIGAGP SYTIDQALNRQEVNVAVDNHYVAILNTHDTSWIIPYIANKDAATAEIKTL WTQISNPKYNDLSLIPETNPELIGTPEEWYGGTSEARDVNVNLAANAIRLTG AKNSRRIPITYAASDKISINGLKIPIPNNDKILISLHAYIPYEMVNGVSSWGS AKDKSLDNLAIYKFKGKGVVIGCEPASVKNKNENORASLAKYVVSAAKRSIA PIWMDNGVSYSGAKDSYGFNRNLTWRWECPIVHALTGLKVKK"
complement(6748..7860)
/gene="CAC0826"
complement(6748..7860)
/gene="CAC0826"
/codon_start=1
/transl_table=11
/product="Endoglucanase family 5"
/protein_id="AAK78802.1"
/db_xref="GI:15023717"
/translation="MFKNLFSKAKISLLTFAFLGTSLSGAYSACAATVPHDMSQOI VNDMKVQNLGNTLDSAPDEGTGNPKTKTKMIDIKIKEAGFNTVPIPTWNSHIGAGP SYTIDQALNRQEVNVAVDNHYVAILNTHDTSWIIPYIANKDAATAEIKTL WTQISNPKYNDLSLIPETNPELIGTPEEWYGGTSEARDVNVNLAANAIRLTG AKNSRRIPITYAASDKISINGLKIPIPNNDKILISLHAYIPYEMVNGVSSWGS AKDKSLDNLAIYKFKGKGVVIGCEPASVKNKNENORASLAKYVVSAAKRSIA PIWMDNGVSYSGAKDSYGFNRNLTWRWECPIVHALTGLKVKK"
complement(6748..7860)
/gene="CAC0826"
complement(6748..7860)
/gene="CAC0826"
/codon_start=1
/transl_table=11
/product="Endoglucanase family 5"
/protein_id="AAK78802.1"
/db_xref="GI:15023717"
/translation="MFKNLFSKAKISLLTFAFLGTSLSGAYSACAATVPHDMSQOI VNDMKVQNLGNTLDSAPDEGTGNPKTKTKMIDIKIKEAGFNTVPIPTWNSHIGAGP SYTIDQALNRQEVNVAVDNHYVAILNTHDTSWIIPYIANKDAATAEIKTL WTQISNPKYNDLSLIPETNPELIGTPEEWYGGTSEARDVNVNLAANAIRLTG AKNSRRIPITYAASDKISINGLKIPIPNNDKILISLHAYIPYEMVNGVSSWGS AKDKSLDNLAIYKFKGKGVVIGCEPASVKNKNENORASLAKYVVSAAKRSIA PIWMDNGVSYSGAKDSYGFNRNLTWRWECPIVHALTGLKVKK"
complement(6748..7860)
/gene="CAC0826"
complement(6748..7860)
/gene="CAC0826"
/codon_start=1
/transl_table=11
/product="Endoglucanase family 5"
/protein_id="AAK78802.1"
/db_xref="GI:15023717"
/translation="MFKNLFSKAKISLLTFAFLGTSLSGAYSACAATVPHDMSQOI VNDMKVQNLGNTLDSAPDEGTGNPKTKTKMIDIKIKEAGFNTVPIPTWNSHIGAGP SYTIDQALNRQEVNVAVDNHYVAILNTHDTSWIIPYIANKDAATAEIKTL WTQISNPKYNDLSLIPETNPELIGTPEEWYGGTSEARDVNVNLAANAIRLTG AKNSRRIPITYAASDKISINGLKIPIPNNDKILISLHAYIPYEMVNGVSSWGS AKDKSLDNLAIYKFKGKGVVIGCEPASVKNKNENORASLAKYVVSAAKRSIA PIWMDNGVSYSGAKDSYGFNRNLTWRWECPIVHALTGLKVKK"
complement(6748..7860)
/gene="CAC0826"
complement(6748..7860)
/gene="CAC0826"
/codon_start=1
/transl_table=11
/product="Endoglucanase family 5"
/protein_id="AAK78802.1"
/db_xref="GI:15023717"
/translation="MFKNLFSKAKISLLTFAFLGTSLSGAYSACAATVPHDMSQOI VNDMKVQNLGNTLDSAPDEGTGNPKTKTKMIDIKIKEAGFNTVPIPTWNSHIGAGP SYTIDQALNRQEVNVAVDNHYVAILNTHDTSWIIPYIANKDAATAEIKTL WTQISNPKYNDLSLIPETNPELIGTPEEWYGGTSEARDVNVNLAANAIRLTG AKNSRRIPITYAASDKISINGLKIPIPNNDKILISLHAYIPYEMVNGVSSWGS AKDKSLDNLAIYKFKGKGVVIGCEPASVKNKNENORASLAKYVVSAAKRSIA PIWMDNGVSYSGAKDSYGFNRNLTWRWECPIVHALTGLKVKK"
complement(6748..7860)
/gene="CAC0826"
complement(6748..7860)
/gene="CAC0826"
/codon_start=1
/transl_table=11
/product="Endoglucanase family 5"
/protein_id="AAK78802.1"
/db_xref="GI:15023717"
/translation="MFKNLFSKAKISLLTFAFLGTSLSGAYSACAATVPHDMSQOI VNDMKVQNLGNTLDSAPDEGTGNPKTKTKMIDIKIKEAGFNTVPIPTWNSHIGAGP SYTIDQALNRQEVNVAVDNHYVAILNTHDTSWIIPYIANKDAATAEIKTL WTQISNPKYNDLSLIPETNPELIGTPEEWYGGTSEARDVNVNLAANAIRLTG AKNSRRIPITYAASDKISINGLKIPIPNNDKILISLHAYIPYEMVNGVSSWGS AKDKSLDNLAIYKFKGKGVVIGCEPASVKNKNENORASLAKYVVSAAKRSIA PIWMDNGVSYSGAKDSYGFNRNLTWRWECPIVHALTGLKVKK"
complement(6748..7860)
/gene="CAC0826"
complement(6748..7860)
/gene="CAC0826"
/codon_start=1
/transl_table=11
/product="Endoglucanase family 5"
/protein_id="AAK78802.1"
/db_xref="GI:15023717"
/translation="MFKNLFSKAKISLLTFAFLGTSLSGAYSACAATVPHDMSQOI VNDMKVQNLGNTLDSAPDEGTGNPKTKTKMIDIKIKEAGFNTVPIPTWNSHIGAGP SYTIDQALNRQEVNVAVDNHYVAILNTHDTSWIIPYIANKDAATAEIKTL WTQISNPKYNDLSLIPETNPELIGTPEEWYGGTSEARDVNVNLAANAIRLTG AKNSRRIPITYAASDKISINGLKIPIPNNDKILISLHAYIPYEMVNGVSSWGS AKDKSLDNLAIYKFKGKGVVIGCEPASVKNKNENORASLAKYVVSAAKRSIA PIWMDNGVSYSGAKDSYGFNRNLTWRWECPIVHALTGLKVKK"
complement(6748..7860)
/gene="CAC0826"
complement(6748..7860)
/gene="CAC0826"
/codon_start=1
/transl_table=11
/product="Endoglucanase family 5"
/protein_id="AAK78802.1"
/db_xref="GI:15023717"
/translation="MFKNLFSKAKISLLTFAFLGTSLSGAYSACAATVPHDMSQOI VNDMKVQNLGNTLDSAPDEGTGNPKTKTKMIDIKIKEAGFNTVPIPTWNSHIGAGP SYTIDQALNRQEVNVAVDNHYVAILNTHDTSWIIPYIANKDAATAEIKTL WTQISNPKYNDLSLIPETNPELIGTPEEWYGGTSEARDVNVNLAANAIRLTG AKNSRRIPITYAASDKISINGLKIPIPNNDKILISLHAYIPYEMVNGVSSWGS AKDKSLDNLAIYKFKGKGVVIGCEPASVKNKNENORASLAKYVVSAAKRSIA PIWMDNGVSYSGAKDSYGFNRNLTWRWECPIVHALTGLKVKK"
complement(6748..7860)
/gene="CAC0826"
complement(6748..7860)
/gene="CAC0826"
/codon_start=1
/transl_table=11
/product="Endoglucanase family 5"
/protein_id="AAK78802.1"
/db_xref="GI:15023717"
/translation="MFKNLFSKAKISLLTFAFLGTSLSGAYSACAATVPHDMSQOI VNDMKVQNLGNTLDSAPDEGTGNPKTKTKMIDIKIKEAGFNTVPIPTWNSHIGAGP SYTIDQALNRQEVNVAVDNHYVAILNTHDTSWIIPYIANKDAATAEIKTL WTQISNPKYNDLSLIPETNPELIGTPEEWYGGTSEARDVNVNLAANAIRLTG AKNSRRIPITYAASDKISINGLKIPIPNNDKILISLHAYIPYEMVNGVSSWGS AKDKSLDNLAIYKFKGKGVVIGCEPASVKNKNENORASLAKYVVSAAKRSIA PIWMDNGVSYSGAKDSYGFNRNLTWRWECPIVHALTGLKVKK"
complement(6748..7860)
/gene="CAC0826"
complement(6748..7860)
/gene="CAC0826"
/codon_start=1
/transl_table=11
/product="Endoglucanase family 5"
/protein_id="AAK78802.1"
/db_xref="GI:15023717"
/translation="MFKNLFSKAKISLLTFAFLGTSLSGAYSACAATVPHDMSQOI VNDMKVQNLGNTLDSAPDEGTGNPKTKTKMIDIKIKEAGFNTVPIPTWNSHIGAGP SYTIDQALNRQEVNVAVDNHYVAILNTHDTSWIIPYIANKDAATAEIKTL WTQISNPKYNDLSLIPETNPELIGTPEEWYGGTSEARDVNVNLAANAIRLTG AKNSRRIPITYAASDKISINGLKIPIPNNDKILISLHAYIPYEMVNGVSSWGS AKDKSLDNLAIYKFKGKGVVIGCEPASVKNKNENORASLAKYVVSAAKRSIA PIWMDNGVSYSGAKDSYGFNRNLTWRWECPIVHALTGLKVKK"
complement(6748..7860)
/gene="CAC0826"
complement(6748..7860)
/gene="CAC0826"
/codon_start=1
/transl_table=11
/product="Endoglucanase family 5"
/protein_id="AAK78802.1"
/db_xref="GI:15023717"
/translation="MFKNLFSKAKISLLTFAFLGTSLSGAYSACAATVPHDMSQOI VNDMKVQNLGNTLDSAPDEGTGNPKTKTKMIDIKIKEAGFNTVPIPTWNSHIGAGP SYTIDQALNRQEVNVAVDNHYVAILNTHDTSWIIPYIANKDAATAEIKTL WTQISNPKYNDLSLIPETNPELIGTPEEWYGGTSEARDVNVNLAANAIRLTG AKNSRRIPITYAASDKISINGLKIPIPNNDKILISLHAYIPYEMVNGVSSWGS AKDKSLDNLAIYKFKGKGVVIGCEPASVKNKNENORASLAKYVVSAAKRSIA PIWMDNGVSYSGAKDSYGFNRNLTWRWECPIVHALTGLKVKK"
complement(6748..7860)
/gene="CAC0826"
complement(6748..7860)
/gene="CAC0826"
/codon_start=1
/transl_table=11
/product="Endoglucanase family 5"
/protein_id="AAK78802.1"
/db_xref="GI:15023717"
/translation="MFKNLFSKAKISLLTFAFLGTSLSGAYSACAATVPHDMSQOI VNDMKVQNLGNTLDSAPDEGTGNPKTKTKMIDIKIKEAGFNTVPIPTWNSHIGAGP SYTIDQALNRQEVNVAVDNHYVAILNTHDTSWIIPYIANKDAATAEIKTL WTQISNPKYNDLSLIPETNPELIGTPEEWYGGTSEARDVNVNLAANAIRLTG AKNSRRIPITYAASDKISINGLKIPIPNNDKILISLHAYIPYEMVNGVSSWGS AKDKSLDNLAIYKFKGKGVVIGCEPASVKNKNENORASLAKYVVSAAKRSIA PIWMDNGVSYSGAKDSYGFNRNLTWRWECPIVHALTGLKVKK"
complement(6748..

complement (8144. .9007) /gene="CAC0827" /codon_start=1 /transl_table=11 /product="Fructose-bisphosphate aldolase" /protein_id="AAK78803.1" /db_xref="GI:15023718" /translation="MLVSAKEMVDKAREGKAVAGQFNINNLEWTKAILLTAQENKSPV ILGVSGAGKYGWGHVAVGMVNGMLQNLITVPVALHLDHGSYEGAFKVIIEAGFSSV MFQSDYIAENIETKEVIGKAANAGLSVAEAGVAGEEDGVVGGVEGAPAECKK IADLGVYDIIAAGIGNIHGVYIPANWAGLRFVLEIDIKNAVGTWPLVLHGHTGIGIPDMIA KATSLGVANINWVTEQQLVFAEATRYKIEEGKQKQKGFDRKLLAFGFEAIRKAYKE KMYLFGSVNRA"	REFERENCE AUTHORS 1 (bases 1 to 12929) Deppenmeier, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R.A., Martinez-Arias, R., Henne, A., Wierzer, A., Baeumer, S., Jacobi, C., Brueggemann, H., Lienard, T., Christmann, A., Boemecke, M., Steckel, S., Bhattacharyya, A., Lykidis, A., Overbeek, R., Klenk, H.-P., Gunsalus, R.P., Fritz, H.-J. and Gottschalk, G. The genome of Methanosarcina mazei: evidence for lateral gene transfer between bacteria and archaea J. Mol. Microbiol. Biotechnol. 4 (4), 453-461 (2002) 22120827 FUEMED 12125824
complement (9320. .10246) /gene="CAC0828" complement (9320. .10246) /gene="CAC0828" /codon_start=1 /transl_table=11 /product="Hypothetical protein" /protein_id="AAK78804.1" /db_xref="GI:15023719"	REFERENCE AUTHORS 2 (bases 1 to 12929) Deppenmeier, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R.A., Martinez-Arias, R., Henne, A., Wierzer, A., Baeumer, S., Jacobi, C., Brueggemann, H., Lienard, T., Christmann, A., Boemecke, M., Steckel, S., Bhattacharyya, A., Lykidis, A., Overbeek, R., Klenk, H.-P., Gunsalus, R.P., Fritz, H.-J. and Gottschalk, G. Direct submission Submitted (23-AUG-2001) Goettingen Genomics Laboratory, Institute for Microbiology and Genetics, Griesbachstrasse 8, Goettingen 37077, Germany
complement (9320. .10246) /gene="CAC0828" complement (9320. .10246) /gene="CAC0828" /codon_start=1 /transl_table=11 /product="Hypothetical protein" /protein_id="AAK78804.1" /db_xref="GI:15023719"	REFERENCE AUTHORS 2 (bases 1 to 12929) Deppenmeier, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R.A., Martinez-Arias, R., Henne, A., Wierzer, A., Baeumer, S., Jacobi, C., Brueggemann, H., Lienard, T., Christmann, A., Boemecke, M., Steckel, S., Bhattacharyya, A., Lykidis, A., Overbeek, R., Klenk, H.-P., Gunsalus, R.P., Fritz, H.-J. and Gottschalk, G. Direct submission Submitted (23-AUG-2001) Goettingen Genomics Laboratory, Institute for Microbiology and Genetics, Griesbachstrasse 8, Goettingen 37077, Germany
complement (9320. .10246) /gene="CAC0828" complement (9320. .10246) /gene="CAC0828" /codon_start=1 /transl_table=11 /product="Hypothetical protein" /protein_id="AAK78804.1" /db_xref="GI:15023719"	REFERENCE AUTHORS 2 (bases 1 to 12929) Deppenmeier, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R.A., Martinez-Arias, R., Henne, A., Wierzer, A., Baeumer, S., Jacobi, C., Brueggemann, H., Lienard, T., Christmann, A., Boemecke, M., Steckel, S., Bhattacharyya, A., Lykidis, A., Overbeek, R., Klenk, H.-P., Gunsalus, R.P., Fritz, H.-J. and Gottschalk, G. Direct submission Submitted (23-AUG-2001) Goettingen Genomics Laboratory, Institute for Microbiology and Genetics, Griesbachstrasse 8, Goettingen 37077, Germany
complement (9320. .10246) /gene="CAC0828" complement (9320. .10246) /gene="CAC0828" /codon_start=1 /transl_table=11 /product="Hypothetical protein" /protein_id="AAK78804.1" /db_xref="GI:15023719"	REFERENCE AUTHORS 2 (bases 1 to 12929) Deppenmeier, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R.A., Martinez-Arias, R., Henne, A., Wierzer, A., Baeumer, S., Jacobi, C., Brueggemann, H., Lienard, T., Christmann, A., Boemecke, M., Steckel, S., Bhattacharyya, A., Lykidis, A., Overbeek, R., Klenk, H.-P., Gunsalus, R.P., Fritz, H.-J. and Gottschalk, G. Direct submission Submitted (23-AUG-2001) Goettingen Genomics Laboratory, Institute for Microbiology and Genetics, Griesbachstrasse 8, Goettingen 37077, Germany
complement (9320. .10246) /gene="CAC0828" complement (9320. .10246) /gene="CAC0828" /codon_start=1 /transl_table=11 /product="Hypothetical protein" /protein_id="AAK78804.1" /db_xref="GI:15023719"	REFERENCE AUTHORS 2 (bases 1 to 12929) Deppenmeier, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R.A., Martinez-Arias, R., Henne, A., Wierzer, A., Baeumer, S., Jacobi, C., Brueggemann, H., Lienard, T., Christmann, A., Boemecke, M., Steckel, S., Bhattacharyya, A., Lykidis, A., Overbeek, R., Klenk, H.-P., Gunsalus, R.P., Fritz, H.-J. and Gottschalk, G. Direct submission Submitted (23-AUG-2001) Goettingen Genomics Laboratory, Institute for Microbiology and Genetics, Griesbachstrasse 8, Goettingen 37077, Germany
complement (9320. .10246) /gene="CAC0828" complement (9320. .10246) /gene="CAC0828" /codon_start=1 /transl_table=11 /product="Hypothetical protein" /protein_id="AAK78804.1" /db_xref="GI:15023719"	REFERENCE AUTHORS 2 (bases 1 to 12929) Deppenmeier, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R.A., Martinez-Arias, R., Henne, A., Wierzer, A., Baeumer, S., Jacobi, C., Brueggemann, H., Lienard, T., Christmann, A., Boemecke, M., Steckel, S., Bhattacharyya, A., Lykidis, A., Overbeek, R., Klenk, H.-P., Gunsalus, R.P., Fritz, H.-J. and Gottschalk, G. Direct submission Submitted (23-AUG-2001) Goettingen Genomics Laboratory, Institute for Microbiology and Genetics, Griesbachstrasse 8, Goettingen 37077, Germany
complement (9320. .10246) /gene="CAC0828" complement (9320. .10246) /gene="CAC0828" /codon_start=1 /transl_table=11 /product="Hypothetical protein" /protein_id="AAK78804.1" /db_xref="GI:15023719"	REFERENCE AUTHORS 2 (bases 1 to 12929) Deppenmeier, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R.A., Martinez-Arias, R., Henne, A., Wierzer, A., Baeumer, S., Jacobi, C., Brueggemann, H., Lienard, T., Christmann, A., Boemecke, M., Steckel, S., Bhattacharyya, A., Lykidis, A., Overbeek, R., Klenk, H.-P., Gunsalus, R.P., Fritz, H.-J. and Gottschalk, G. Direct submission Submitted (23-AUG-2001) Goettingen Genomics Laboratory, Institute for Microbiology and Genetics, Griesbachstrasse 8, Goettingen 37077, Germany
complement (9320. .10246) /gene="CAC0828" complement (9320. .10246) /gene="CAC0828" /codon_start=1 /transl_table=11 /product="Hypothetical protein" /protein_id="AAK78804.1" /db_xref="GI:15023719"	REFERENCE AUTHORS 2 (bases 1 to 12929) Deppenmeier, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R.A., Martinez-Arias, R., Henne, A., Wierzer, A., Baeumer, S., Jacobi, C., Brueggemann, H., Lienard, T., Christmann, A., Boemecke, M., Steckel, S., Bhattacharyya, A., Lykidis, A., Overbeek, R., Klenk, H.-P., Gunsalus, R.P., Fritz, H.-J. and Gottschalk, G. Direct submission Submitted (23-AUG-2001) Goettingen Genomics Laboratory, Institute for Microbiology and Genetics, Griesbachstrasse 8, Goettingen 37077, Germany
complement (9320. .10246) /gene="CAC0828" complement (9320. .10246) /gene="CAC0828" /codon_start=1 /transl_table=11 /product="Hypothetical protein" /protein_id="AAK78804.1" /db_xref="GI:15023719"	REFERENCE AUTHORS 2 (bases 1 to 12929) Deppenmeier, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R.A., Martinez-Arias, R., Henne, A., Wierzer, A., Baeumer, S., Jacobi, C., Brueggemann, H., Lienard, T., Christmann, A., Boemecke, M., Steckel, S., Bhattacharyya, A., Lykidis, A., Overbeek, R., Klenk, H.-P., Gunsalus, R.P., Fritz, H.-J. and Gottschalk, G. Direct submission Submitted (23-AUG-2001) Goettingen Genomics Laboratory, Institute for Microbiology and Genetics, Griesbachstrasse 8, Goettingen 37077, Germany
complement (9320. .10246) /gene="CAC0828" complement (9320. .10246) /gene="CAC0828" /codon_start=1 /transl_table=11 /product="Hypothetical protein" /protein_id="AAK78804.1" /db_xref="GI:15023719"	REFERENCE AUTHORS 2 (bases 1 to 12929) Deppenmeier, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R.A., Martinez-Arias, R., Henne, A., Wierzer, A., Baeumer, S., Jacobi, C., Brueggemann, H., Lienard, T., Christmann, A., Boemecke, M., Steckel, S., Bhattacharyya, A., Lykidis, A., Overbeek, R., Klenk, H.-P., Gunsalus, R.P., Fritz, H.-J. and Gottschalk, G. Direct submission Submitted (23-AUG-2001) Goettingen Genomics Laboratory, Institute for Microbiology and Genetics, Griesbachstrasse 8, Goettingen 37077, Germany
complement (9320. .10246) /gene="CAC0828" complement (9320. .10246)<	

```
/db_xref="GI:20906101"
/translation="MSENSENAPEKVEFTPIGYVENDYFKPEYKEEVEYKSVILVRE
LDVGLVYVEELKYLIFYSKSGYKLIHRRYDGMGSGVFAERSPYRPNIGLTL
VELLVKGVNLYVYKGLDAINGTVPVLDIKPYMKDLSEKENTK"
gene
complement (5335..5853)
/genes="MM1604"
CDS
complement (5335..5853)
/codon_start=1
/genes="MM1604"
/transl_table=11
/product="tungsten formylmethanofuran dehydrogenase,
subunit E"
/protein_id="AA031300.1"
/db_xref="GI:20906102"
/translation="MAVLAERFKORSEDELVAVENRSCAVDAIQAINGCTCGKN
LIPKEGKVYFYKSGSKALRISLKPDALQDSRHRALAKRAGTASPEERKEFQ
ASHDASKRILENPESELFRVSEVNIPEPEKAIITPTVICKSGEGFPELGRVKNGE
IVCISCFEAKNE"
gene
7350..8138
/genes="MM1605"
CDS
7350..8138
/genes="MM1605"
/codon_start=1
/transl_table=11
/product="conserved protein"
/protein_id="AA031301.1"
/db_xref="GI:20906103"
/translation="MNSLLELIIFLSEKRNLLFLKEGPKNIEIKTNLDTSVAAIL
PQIKRDSLVFRFDSYFLGPLGIAVAERHMAVRVNVFSRNYDYADHAEVCLP
SPUKSRIGLENCIIEPDRDTHLPBHEFVEKLTGVRKIKGISAIFPLPSLPLS
FANGADVLLVTLVPVPEIKGEVSELRELLDNTCLYCEKIKBESFHVTDKPLS
LSLPSNGAFDTHRDVLCFDPVALKAGDLFAYINDRSEKIKV"
gene
8236..9159
/genes="MM1606"
CDS
8236..9159
/genes="MM1606"
/notes="aIF-2B2"
/codon_start=1
/transl_table=11
/product="Archaeal protein translation initiation factor
2B subunit 2"
/protein_id="AA031302.1"
/db_xref="GI:20906104"
/translation="MKEVQETAEKIRMEIRGAGRIAKAAAEVIRDVAAAGLEASIEE
FAGITEVNYLISRTPAVSNAPVNLKSSRYSDNVENARKEVINNANLFIERADKA
LGRIGKAGRVQDGDVIMTHCNHSAALSIIITAFEDGKNTSVIATESRPRROGLTV
RLNGPGIPTTLVDSAVRYVMKDYKIVGADALANGALVNIKIGTSOLALAHAR
KSPVAAETFPKSPSTILGIPIEIERSTDEVDPVSLAELPHVQVKNPAFDTPSEY
IDMIVDIGIIPPAVAYIVIKHGWEEFEI"
gene
9219..9683
/genes="MM1607"
CDS
9219..9683
/genes="MM1607"
/codon_start=1
/transl_table=11
/product="Acetyltransferase"
/protein_id="AA031303.1"
/db_xref="GI:20906105"
/translation="MINVEYTDGAQBSLYLIESLWQSRTHRNKSKYPADTYANKRF
QDRVELTDDSAARVNLVKDQGTGIGYICITIDKEMTIGDSLYLVEYRKHGI
GSLWEKALEWLDKKNKSKIVAVGDGNEKVIDFNYHGHIRKILIEQVSN"
gene
complement (9752..10462)
/genes="MM1608"
CDS
complement (9752..10462)
/genes="MM1608"
/codon_start=1
/transl_table=11
/product="ABC transporter, ATP-binding protein"
/protein_id="AA031304.1"
/db_xref="GI:20906106"
/translation="MREEKEGTAGRKVIMQIENLSKTYMQGKI PVHARCACTVR
KGLLALMPGSGSKSTLALIGLEKATGKIILDTGLTSLVPEKLLPVREKRGIF
IQHYNLFTLSALENVELAMFRBQKKEKERSKLELDEGLDEHDKHPTLSGG
```

```
EQQRVSIARALANKPALILADEPTGEVDSKTRDSVWQIFKELSGKQTILVWTHDPEV
AKECVRVIRITDGMKIH"
complement (10455..11570)
/genes="MM1609"
CDS
complement (10455..11570)
/genes="MM1609"
/codon_start=1
/transl_table=11
/product="ABC transporter, permease protein"
/protein_id="AA031305.1"
/db_xref="GI:20906107"
/translation="WLDLIIRNITNRKIRGALTICGIALGIFAVIVMGAMSENPHOTF
ERSMSTDKIRVFAESGVFGGLTEKVDLVRVAGVADAYGLLMTPTPDRMGWTG
KQLGVPEKSVFNVPOLKAGFLDPGDSYSVWGNNTKRNLEIGSKFHIHDKY
FTVIGLIDYTSIFDNVILILETAQLIYSGNSVSIIPAPDPRVDAMLSKRIBLS
VAGTILSPGLEVQARQSFVSVITISSGLAAIIGGCVWNTMLMSVAERISFLS
ILKAIGATKDIILLTGLSEASLMGLGLGIIVGIGAVQIMNANLATTIRVILFLITP
RLLIAMFLALLIGALSGLPAYRASKVSPMEALKHA"
gene
complement (11626..12813)
/genes="MM1610"
CDS
complement (11626..12813)
/genes="MM1610"
/codon_start=1
/transl_table=11
/product="hypothetical binding protein of ABC transporter"
/protein_id="AA031306.1"
/db_xref="GI:20906108"
/translation="MSLPKKNCKLREFLSLFLIFFCTOINPVYGVANGAA
TEENENRIIGFSDHLSDATVRPEPHENSVLFTLSGKETIKSETFHLVSVKK
DOELSKVFWLEEDFKDRDSYTAOLFKVEGNQILETOKTSPSYNHLNKKVDF
SADSEKSFLLDVSANLILVQMPERGIIVDFDLKLLSKDIYSSSLKNPVTDAY
ATXWPLLELDRSYTALLKHSHSPELITAYMSEFMAEKVILSDSDINVDYASV
TVTGRSQVFNFIIVRVLPPEGRVRFETADILTAGKEDTVGIIWQGVPRGDYNVK
IDVNLDDGEILDSHETVLRVFERPYAELNPEETPALSSQAATGILLCVAVFLGRKRG
KKD"
ORIGIN
Query Match 13.5%; Score 82; DB 1; Length 12929;
Best Local Similarity 50.8%; Pred. No. 9.3e-14;
Matches 261; Conservative 0; Mismatches 235; Indels 18; Gaps 2;
QY 78 AGGACAAATGACGACACTGACGAGCGTCGGTTACAGGAAATCGACTTTGTCAACTG 137
DB 10328 AGGAGAGCTCCTCGTATATATGGGACCTTCGGTTACAGGAAATCAACCTTCTGGCTCT 10269
QY 138 TCTTGGCACATTCACAAACCACTTCGGGACAGATCCTTGTGAGGATGTAGACCTTCT 197
DB 10268 CATAGGCACATTTGAAAAGCCCTGACGAAATAATTAATTTCTGACGGGACAGCTTAC 10209
QY 198 GAAACTCTCTACGCTAAGCAACGGTTATACAGGAAATAATACGGTGGCTATTATTCTCA 257
DB 10208 ATCAGTACCTGAAAAGCTGCTCCCGGCGTGAGAGAGAAAAATAGGCTTTATTTTCA 10149
QY 258 AGATTATGCTTGTATCCCGACAGACAGTTAAATTCACCTTCAGCTTCGGTGG --- 313
DB 10148 GCATACAACTGATTCCTACATTTGAGCCTTGAACCTTGAACCTTGCATGGCTT 10089
QY 314 -----AAAAACAAATGGCTGAAATTCCTCAAGTACTTCATGCTG ---TTGG 359
DB 10088 TTCAAGATGCAAAAAAGAGAGAAAAAGGCAAAATCCCTTCTTGAAGACCTGGG 10029
QY 360 TCTTGAGTCGTTCCAGGAAAGCCAGTTTTTGAACCTCTCTGTTGGCGGACACCAACGAAC 419
DB 10028 GCTGAAGACCGGCTCAATCAATAAAGCTCTCAGGTGGAGAACGCAAGAGGT 9969
QY 420 TCGGTTGCCCGGTACTGTCTAAAAATCCCGAAATAATTTCTGGCTGATGAACCAACCG 479
DB 9968 TTCTATTGCAAGGCTCTGGCAATAAATCCCTGCCCTGATTTCTGGCTGACGAGCTACAG 9909
QY 480 AGCTCTAGATTACAAACAGTGTAGCTAGTCATAGAGGATTCAGAGAGACTCGCCGACAA 539
DB 9908 TGAGTGCATAGCAAAACAGGAGACTCCGTTGTGCGAGATTTTAAAGAAATTGAGCCAG 9849
```

[illegible]

APETLPGDTLLKKGKDEMGVYVYRSMTHAFNGKDGVGSGKMTTGEKGTSEBTGPV
GLYKNTNPPLEKTNWDIDPTGLRIGLRLITNRYKLPIMITENGELGEYDSLTDHDK
IHDEYRLRYRAHALAQERAITDGVEMLGCTWSTFDLLSWLNGYQKRYGVFVYDRE
NDEKRLRYKXKDSFYWKYKTIENGANLVBEOQK"

9935..10237

gene

/gene="lin2905"

CDS

9935..10237

/gene="lin2905"

/note="similar to PTS cellobiose-specific enzyme IIB"

/codon_start=1

/transl_table=11

/protein_id="CAC98130.1"

/db_xref="GI:16415446"

/db_xref="GOA:Q926Y5"

/db_xref="SPTREMBL:Q926Y5"

/translation="MKILVLCAGAGSTSLVTKMAHATPSIGSEIEIHALPVSEASS
VDRMDIVMUGPQRYQKQPVDELVQGRIPVIVDMKDYGNLNGKAVLEKAFABIG"

10258..11623

gene

/gene="lin2906"

RBS

10258..10263

/gene="lin2906"

10271..11623

/gene="lin2906"

/note="similar to PTS cellobiose-specific enzyme IIC"

/codon_start=1

/transl_table=11

/protein_id="CAC98131.1"

/db_xref="GI:16415447"

/db_xref="GOA:Q926Y4"

/db_xref="SPTREMBL:Q926Y4"

/translation="MSIMSKFEHGMERVLVPVANKLSQRHIAAIRDAFLLVPLIMA
GSIITLINFALVSDGFIKILFLGKIFPNLADQAVFSPVMOGNTIMAILIVFLVA
RNLAIFKQDRLCLGTSIGAFVITYPTVVDNASTYITKFLGAQSLFVAITVAIT
GEVSRILARSLKMPDQPPAVARSFKVLPVITILFVSINYLIIAPEGIL
DLVVTQAPLKDGTNVFVSVIIIGLVNLLWGLHGPNVAAIRDTIETPNLNL
SVYAHGASGAPVPATWAGLNDGFANYGSGWTLGLLIALFIASRRADYRDIAKSL
APGFENEPVIFGLPVLPIVPIVPIIIPAINLIGYEFISTKLIPPVAYQVPTT
FGPLIFPGTGGNALLVGLLCLAVATVILPVLVSNKIAASDAAMKRNASTEQ"

11635..12519

gene

/gene="lin2907"

CDS

11635..12519

/gene="lin2907"

/note="similar to xylose operon regulatory protein and to

glucose kinase"

/codon_start=1

/transl_table=11

/protein_id="CAC98132.1"

/db_xref="GI:16415448"

/db_xref="GOA:Q926Y3"

/db_xref="SPTREMBL:Q926Y3"

/translation="MKIAAFDGTALMGVLPVPHGSEIILTKSAEIIASDGDQILAE
KLFAENTDVTVIAPVPGVYVNPKTGLITMGAIIRDENLNKLEWLEATGGLPAIEN
DANCALAEKWLGGQDLDFCLTIGTIGTGIFSGNALVGRGFRFAGFYMFSE
PCAFPGKYLNETTLMVLRQVAILTGRPLKEITGEIEIFANYADHDPISERLINEF
VTGCTGLNLIYLPDPIHPIFGGITSRPTFTELKHMVASFGLRDTIETATKNO
AGLGAUVHLEQENRHE"

12497..12502

gene

/gene="lin2908"

RBS

12497..12502

/gene="lin2908"

Query Match 13.3%; Score 80.8; DB 1; Length 96208;
Best Local Similarity 47.2%; Pred. No. 3.7e-13;

Matches 284; Conservative 0; Mismatches 312; Indels 6; Gaps 1;

QY 1 ATGATAGAAATCAATGACCTCAAGAAATCTTTTGGCGGTTCGATCTTATGCGAAGGTCTC 60

DB 19809 ATGATTGAATTAGITTAATGTCAGTAAAGAACATACAGATAAATGATTTGGAAAAAGTC 19750

QY 61 AGTCATAGTTTATACCAAGGACATACACACTGACTGAGCGTCCGGTTCAGGAAAA 120

DB 19749 TCTCTTTCTATGGAGCTGGCGAATTTATTGCACTGCTCGCGAGAGTGTAGTGGGAAG 19690

QY 121 TCGACTTTTGCTCAACTGTCTTGGCACACTTGACAAACCAAGTTCGGGACAGATCCTTGTG 180
DB 19689 ACAACGCTGCTAAATGTTATCGGACACCTAGATTCAAAAGATAGTGGCAAGTATTATT 19630
QY 181 GAGGATGTAGACCTTCTGAAACTCTCTACGGGTGAACACCGGTATATACAGGAAAAATACG 240
DB 19629 AACGAGATGGAATATCAGA-----CGAAAAAAGAGGTTATGACTCTTAAAAAAGAGGTG 19576
QY 241 GTGGCTATTATTTCAAGATTATGCTTGTATCCCGACAGACAGATTAAATTCACACCTT 300
DB 19575 TTAGGTTTATATCTCCAAATATCTATGTATGGAATGAACAGTGTAGAAAACTTA 19516
QY 301 CAGCTTGGCGTGGAAAAACACAAATGGCTGAAATTCCTCAAGTACTTTCATGCTGTGGT 360
DB 19515 TCCATTACAGGTGGGAAAAATCGCAAGCTGATGATAGAGCATTTTGAAGAAGTAGGAATG 19456
QY 361 CTGTAGTGGTTCGAGAAAAAGCCAGTTTGTGAACTCTCTGGTGGCGGACACACACGAACT 420
DB 19455 GATGAGAGCTATTTAGCAAAAAAAGTATACCAATTAAGTGGTGGAGAAAAACACGGATT 19396
QY 421 GCGTTGGCGCGGTACTGCTCAAAAATCCCGAATAATCTGGCTGATGAACCAACCGGA 480
DB 19395 GCTATTGTGGCATTTTACTCAACCATTTTCAACTTTTACTTGGGACGAACCAACTGGC 19336
QY 481 GCTCTAGATTAAACAAAGTGTAGTCTCATAGAGCATTTGAGAGCACTGCGCGACAAA 540
DB 19335 AATTAGATGATAAAAAACAAAAATCATTTGAATTAATTTCTAGCCTTGAAAAAGCAA 19276
QY 541 GCGGCCACCGTTGTTGCTACGCACTCGCCCTCTCCGAGAAATCAGCGGATACCAATT 600
DB 19275 GGTAGACTATCGTTGTGTGTCAGCATACCGGAAATATCAGGAAAGCAGATCGGTC 19216
QY 601 AT 602
DB 19215 AT 19214

RESULT 10

AX417049/c

LOCUS

Sequence 4040 from Patent WO0228891.

AX417049

VERSION

AX417049.1 GI:21449659

KEYWORDS

SOURCE

Listeria innocua

ORGANISM

Listeria innocua

Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.

REFERENCE

1

AUTHORS

Kunst, F. and Glaser, P.

TITLE

Listeria innocua, genome and applications

JOURNAL

Patent: WO 0228891-A 4040 11-APR-2002;

INSTITUT PASTEUR (FR); CENTRE NATIONAL DE LA RECHERCHE

SCIENTIFIQUE (CNRS) (FR)

LOCATION/Qualifiers

1..311208

/organism="Listeria innocua"

/mol_type="unassigned DNA"

/db_xref="taxon:1642"

/note="seq 2058, original length: 3,011,208 replaced

by-seq 2058: 0.000.001 to 0.349.980-seq 4032: 0.300.001 to

0.649.980-seq 4033: 0.600.001 to 0.949.980-seq 4034:

0.900.001 to 1.249.980-seq 4035: 1.200.001 to

1.549.980-seq 4036: 1.500.001 to 1.849.980-seq 4037:

1.800.001 to 2.149.980-seq 4038: 2.100.001 to

2.449.980-seq 4039: 2.400.001 to 2.749.980-seq 4040:

2.700.001 to 3.049.980"

ORIGIN

Query Match

13.3%; Score 80.8; DB 6; Length 311208;

Best Local Similarity 47.2%; Pred. NO. 5e-13;

Matches 284; Conservative 0; Mismatches 312; Indels 6; Gaps 1;


```
QY 1 ATGATAGAAATCAATGACCTCAAGAAATCTTTGGCGTTCGGATCTTATGGCAAGTCTC 60
Db 234809 ATGATTCGAATAGTTATGTCAGTAAAGATCAAGATAAATGATTTTGGAAAAAGTC 234750
QY 61 AGTCATAAGTTTTTACAGGAACAATGACAGCACTGACTGGAGCGTCGGTTTCAGGAAAA 120
Db 234749 TCTCTTTCTATTGGAGCTGGCGAATTTATTTCAGTCTGTCGGCGAGAGTGTAGTGGGAG 234690
QY 121 TCGACTTTGCTCACTGCTTGGCACACTTGACAAACCAAGTTCCGGACAGATCCTTGTC 180
Db 234689 ACAACGCTGCTAAATGTTATCGGACACCTAGATTCAAAAGATAGTGGGCAAGTTATT 234630
QY 181 GAGGATGTAGACCTTCTGAAACTCTCTACCGCTGAAGCAACGGTTATATACAGGAAAAATACG 240
Db 234629 AACGAGATGAATATCAGA-----CGAAAAAGAGGTTATGACTCTAAAAAAGAGGTG 234576
QY 241 GTGGGCTATTATTTCAGAGTTATGCTTGAATTCCTGATCCCGACAGGACAGTTAAATTCACCTT 300
Db 234575 TTAGGTTTATATTCGAAAAATTTATCTATTGATGGAATGAAACAGTGTAGAAAACTTA 234516
QY 301 CAGCTTCGGTGGAAAAACAAATGCTTCAAGTCTTCAAGTACTTTCATGCTGTTGGT 360
Db 234515 TCCATTACAGTGGGGAATTCGAACTGATGATAGAGCATTTTGAAGAAGATGGAATG 234456
QY 361 CTGAGTCTGTCGAGGAAAAAGCCAGTTTGTAACTCTCTGGTGGCGAAACAAACGAACT 420
Db 234455 GATGAGAGCTATTATAGCAAAAAAAGTATACCAATTAAGTGGTGGAGAAAAACAAACGATT 234396
QY 421 GCGTTGCGCGGCTACTGCTCAAAATCCCGATTAATTCGCTGATGAACCAACCGA 480
Db 234395 GCTATTGTGCGCATTTTACTCAAAACCAATTCACCTTTTCTGCGACGAAACCACTGGC 234336
QY 481 GCTCTAGATTAAACACAGTAGCTAGTCTATAGAAGCATTTGAGACACTTCGCCGCAAA 540
Db 234335 AATTAGATGATAAAAAACAAACAAAAATCATTTGAATTTATCTAGCTTTGAAAAAGCA 234276
QY 541 GCGCCACCGTTGTTGTTGTAGTACTCGCCCTCTTCGGAGATTCAGCGGATACCATT 600
Db 234275 GGTAAAGTACTGCTTTGTGTCACGCAATGACCCGGAATATCAGGAAAAAGCATGGGTC 234216
QY 601 AT 602
Db 234215 AT 234214
```

```
RESULT 11
AX413018 349980 bp DNA linear PAT 02-SEP-2002
LOCUS AX413018
DEFINITION Sequence 9 from Patent WO0228891.
ACCESSION AX413018
VERSION AX413018.1 GI:21445476
KEYWORDS
SOURCE Listeria innocua
ORGANISM Listeria innocua
Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
```

```
REFERENCE
1
AUTHORS Kunst,F. and Glaser,P.
TITLE Listeria innocua, genome and applications
PATENT: WO 0228891-A 9 11-APR-2002;
JOURNAL INSTITUT PASTEUR (FR) ; CENTRE NATIONAL DE LA RECHERCHE
SCIENTIFIQUE (CNRS) (FR)
```

```
FEATURES
source
1..349980
/organism="Listeria innocua"
/mol_type="unassigned DNA"
/db_xref="taxon:1642"
/notes="old seq 9 (684707) splitted in 3-new seq 9 = old
seq 9 from 1 to 349980-seq 4027 = old seq 9 from 300000 to
649980-seq 4028 = old seq 9 from 600000 to 684707"
```

ORIGIN

Query Match 13.3%; Score 80.8; DB 6; Length 349980;
Best Local Similarity 47.2%; Pred. No. 5.le-13;

```
Matches 284; Conservative 0; Mismatches 312; Indels 6; Gaps 1;
QY 1 ATGATAGAAATCAATGACCTCAAGAAATCTTTGGCGTTCGGATCTTATGGCAAGTCTC 60
Db 346005 ATGATTGAATAGTTATGTCAGTAAAGATCAAGATAAATGATTTTGGAAAAAGTC 346064
QY 61 AGTCATAAGTTTTTACAGGAACAATGACAGCACTGACTGGAGCGTCCGGTTTCAGGAAAA 120
Db 346065 TCTCTTTCTATTGGAGCTGGCGAATTTATTTCAGTCTGTCGGCGAGAGTGTAGTGGGAG 346124
QY 121 TCGACTTTGCTCACTGCTTGGCACACTTGACAAACCAAGTTCCGGACAGATCCTTGTC 180
Db 346125 ACAACGCTGCTAAATGTTATCGGACACCTAGATTCAAAAGATAGTGGGCAAGTTATT 346184
QY 181 GAGGATGTAGACCTTCTGAAACTCTCTACCGCTGAAGCAACGGTTATATACAGGAAAAATACG 240
Db 346185 AACGAGATGAATATCAGA-----CGAAAAAGAGGTTATGACTCTAAAAAAGAGGTG 346238
QY 241 GTGGGCTATTATTTCAGAGTTATGCTTGAATTCCTGATCCCGACAGGACAGTTAAATTCACCTT 300
Db 346239 TTAGGTTTATATTCGAAAAATTTATCTATTGATGGAATGAAACAGTGTAGAAAACTTA 346298
QY 301 CAGCTTCGGTGGAAAAACAAATGCTTGAATTCCTCAAGTACTTTCATGCTGTTGGT 360
Db 346299 TCCATTACAGTGGGGAATTCGCAAGCTGATGATAGACATTTTGGAGAAGTAGGATG 346358
QY 361 CTGAGTCTGTCGAGGAAAAAGCCAGTTTGTAACTCTCTGGTGGCGAAACAAACGAACT 420
Db 346359 GATGAGAGCTATTATAGCAAAAAAAGTATACCAATTAAGTGGTGGAGAAAAACAAACGATT 346418
QY 421 GCGTTGCGCGGCTACTGCTCAAAATCCCGAATTAATTCGCTGATGAACCAACCGGA 480
Db 346419 GCTATTGTGCGCATTTTACTCAAAACCAATTCACCTTTTACTTTCGGGACGAACTGGC 346478
QY 481 GCTCTAGATTAAACACAGTAGCTAGTCTATAGAAGCATTTGAGACACTTCGCCGCAAA 540
Db 346479 AATTAGATGATAAAAAACAAACAAAAATCATTTGAATTTTCTAGCTTTGAAAAAGCAA 346538
QY 541 GCGCCACCGTTGTTGTTGTAGTACTCGCCCTCTTCGGAGATTCAGCGGATACCATT 600
Db 346539 GGTAAAGTACTGCTTTGTGTCACGCAATGACCCGGAATATCAGGAAAAAGCATCGGTC 346598
QY 601 AT 602
Db 346599 AT 346600
```

```
RESULT 12
AX417036 349980 bp DNA linear PAT 02-SEP-2002
LOCUS AX417036
DEFINITION Sequence 4027 from Patent WO0228891.
ACCESSION AX417036
VERSION AX417036.1 GI:21449646
KEYWORDS
SOURCE Listeria innocua
ORGANISM Listeria innocua
Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
```

```
REFERENCE
1
AUTHORS Kunst,F. and Glaser,P.
TITLE Listeria innocua, genome and applications
PATENT: WO 0228891-A 4027 11-APR-2002;
JOURNAL INSTITUT PASTEUR (FR) ; CENTRE NATIONAL DE LA RECHERCHE
SCIENTIFIQUE (CNRS) (FR)
```

```
FEATURES
source
1..349980
/organism="Listeria innocua"
/mol_type="unassigned DNA"
/db_xref="taxon:1642"
/notes="seq 9 splitted in 3 part.-seq 9: 1 to 349980-seq
4027: 300001 to 649980-seq 4028: 600001 to 684707"
```

ORIGIN

Query Match 13.3%; Score 80.8; DB 6; Length 349980;

YEAFAKAYATHKDCVPLVDTYDLRVGPSALVAREFGDKINFGQVRIDSGDMAYIS
KKVROOLDAGSTQDKIYASNDLNTILSLAKQAKADIVMGVGTKLITAYDOPALGA
VYKLWMEDESGMDQTKLSNAEKVSTPGKQWRLTKKADGSEGDYITEPSERP
DVARELYMHPDITYINKNVRPNANPLKDIYFVGELIYGLPSLQEIKEFARENLEE
LWDEYKRDLPNPOYPVDDLSQELNWHKORLISQVITVSNTRLDDDE"
/gene="yilag"
/note="synonym: L108967"
4369. .4857
/gene="yilag"
/note="synonym: L108967"
4369. .4857
/gene="yilag"
/note="EVIDENCE BY HOMOLGY BIO15.01 HYPOTHETICAL"
/codon_start=1
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="AAK05199.1"
/db_xref="GI:12724058"
/translation="NMRIERIDLCDARLFDVQKOLDTESKFMLEPBERRRDFSKRTI
GRISRDFVLGADVCKGVGFLSAGKGNARINRTAYVYVGLIEBPHQIGQSOLFEE
LWKEAKAVKRLVLMINNYAALFALFHQFTIEGVREAVCVGDKIIDEFYMSKI
LD"
4918. .5742
/gene="nadE"
/note="synonym: L0203"
4918. .5742
/gene="nadE"
/note="EVIDENCE BY HOMOLGY BIO02.12 Pyridine nucleotides.
DE NOVO BIOSYNTHESIS OF NAD"
/codon_start=1
/transl_table=11
/product="NAD-synthetase (EC 6.3.5.1)"
/protein_id="AAK05200.1"
/db_xref="GI:12724059"
/translation="MTLQDEIEMKELGVKVIDPEIRVSVFLKDYLLKKYPIKSFV
LGISGQDSLAGRLAQIAIEEMRQSTADETYKFAIRLIFYGVQADEDAQALAFIQ
PDVSLVNTKAAVGVQALNEAGIEVDFNFKRIMIKAROMITQAVAGQYQGVILGT
DHAENITGFTFFGFGGADLLPLFLNKRQKALLAELGADPAIYKVPYTDLEEGK
PGLADIALGVTYNDIDYTEGKVISADAKAKIEMWKKTKHKLPISVFDFWK"
5870. .6571
/gene="yiba"
/note="synonym: L110431"
5870. .6571
/gene="yiba"
/note="EVIDENCE BY HOMOLGY BIO13.01 TRANSPORT AND BINDING
PROTEINS General. There are 60 paralogs in L.lactis."
/codon_start=1
/transl_table=11
/product="ABC transporter ATP-binding protein"
/protein_id="AAK05201.1"
/db_xref="GI:12724060"
/translation="MTPIEVNESKIYQSGCEKIVANDOVNPEIEKGLAVILGSSGA
KSTVNLILGMDTNEGEVI IDEBKISFNKELITYRYAVGVQFQVFNWNLNLTGA
LENVELASEIVENALDAKEVLKSVGLEHRLHNPFSQISGGEQORVALAALAKQPKIL
LCDEPTGALDYHTGKILKILQDMARKEGKTVIIVTHNMAIAFIANRVIQMDHGKIQ
IDNNSQPEDIDRIEW"
6565. .9255
/gene="yiba"
/note="synonym: L111162"
6565. .9255
/gene="yiba"
/note="EVIDENCE BY HOMOLGY BIO13.01 TRANSPORT AND BINDING
PROTEINS General"
/codon_start=1
/transl_table=11
/product="ABC transporter permease protein"
/protein_id="AAK05202.1"
/db_xref="GI:12724061"
/translation="MWSYAKKTLNKEIRITSGLSRPFISFLMLLGTFAFVLKLVIS
GPNMRRTAEDFYAGHLADLTSLTGLDSDQQLINETKGVKKAEGFYQDLIVIKGK
ENSLRLFSKDESLTSLMSGLPKQDEIADLYDQYKIQGTIDFTPPKSKDGL
IKNHSFKIVGVKSVVYDGFSTVGTGKNGVALYVTKFAFDSDVVMYIARLSYKN
LQNTLIPDSYVSLTEQKLTENTFKNQPEKLAALKTGPKEIQNEAKSOIVEEENQ
LTQENQLIAQKNOIGNSAQAIEQINACQNNQDCKEXIAKAKELAKQETALQOL
EKPTYQINLUKEGNPGYKTLFDDSTRIDSLSNIFPVVILFAILLVSLTWTWRFEER

GNLGLLKALGYSNRDIRKFMVYGLVSSGALVGLTIIIGHTFLPIAVFNAYTASSTFS
NLRITSLPTIVAFALAIACSLIPAYVVMELKEVPASFLAKPKAGSRILLLEKI
NFIWKRSMFTYKVTARNLFYKRRMLMTIFGVAGCTALLVMGFIKRSISGLSNKQFG
QILHYDMITKKNVNDKEEEDKELASSEIENYLPIAFENLTKEASGLKEQVNL
IATNRSDDLKVTSLKSKNSOKLELNSGAVLSEKFAELLDLKVGSLTLKDSENOP
HKIKVAALTEMVGHYIPNQSVYQVFKDFVANAQILKIKNASTNNIKMSAQFNQ
EDGICVQNSDLKLVIESFMIGLNSVMVLIISCAILGALVVIINLNNINVSREIL
ATIKVLGFYREVLTIIYRETIILSLFGLVGFGLGDFYHQVIMNQLSADQIMPAPGL
LWNLILLSAATFATTLALLAVVHFLKAYDMLGALKSVDP"
complement (9331. .10116)
/gene="cobQ"
/note="synonym: L113931"
complement (9331. .10116)
/gene="cobQ"
/note="EVIDENCE BY HOMOLGY BIO02.09 Riboflavin and
cobalamin. 43% identical to cobyric acid synthase CobQ of
Heliobacillus mobilis, cobalamin biosynthesis pathway"
/codon_start=1
/transl_table=11
/product="cobyric acid synthase"
/protein_id="AAK05203.1"
/db_xref="GI:12724062"
/translation="MAYISLKSNLNPKYSLNVAHLYGDLMTYTGDNGLNMLKYVGE
KLGAEMTFNIVSLGDTYPAKEDYDILVFWGGQDYEQEIADQSLIDLAPLKDYTESK
PLLAICGGYQMLGOVYVNSEGTQIOGTGILGHVYENLRTDRFIGDIETHNEEFQETYY
GFENHSITYLSDREKPLGTVVVGNNPDQTEGLYKNTFGYFHPILSRNARLA
IRLVITTAIYKIGIEIQLPAPEEILADEDKQKQIGDLKRVKEK"
complement (10117. .11466)
/gene="yiba"
/note="synonym: L114717"
complement (10117. .11466)
Query Match 13.2%; Score 80.6; DB 1; Length 11551;
Best Local Similarity 48.9%; Pred. No. 2.5e-13;
Matches 246; Conservative 0; Mismatches 254; Indels 3; Gaps 1;
QY 59 TCAGTCATAAGTGTTCACGAGAACATGACAGCAGCTGACTGGAGCGTCGCGTT CAGGAA 118
DB 5946 TAAATTTTGAATTTGAAAAGAGGAGAAATTGGCTGTCTATTTTAGTTCTCTCGGGCAGGTA 6005
QY 119 AATCAGCTTGTCTCACTGTCTTGGCACACTTGACAAACCAAGTTCGGACAGATCCCTTG 178
DB 6006 AATCAACAGTCTCTAATAATTTTAGGAGGAATGGATACAAATGATGAGGAGAGGTTATTA 6065
QY 179 TCGAGGATGTAGACCTTCTGAAACTCTCTACCGTAAAGCAACGGTTATACAGGAAAAATA 238
DB 6066 TTGATGAGAAAAAATCTCTAATTTTCAACAAGAAATTCATTACCTACCGAGCTTATG 6125
QY 239 CGGTGGGTATTATTTCAAGATATATGCTTGTATCCCGACAGACAGTAAATTCACAC 298
DB 6126 CGGTGGTTCGTTTTCATATTTTATATCTTGTCAATAATTGACAGCAGCTAGAAAAATG 6185
QY 299 TTCAGCTTCGGTGAGAAAAACACAAATGGCTGAAATTCCTC--AAGTACTTTCATGCTG 355
DB 6186 TCGAATTAGCTTCTGAAATTTGAAATGCTCTTGATGCTTAAGAAGTGTCTGAGAGTG 6245
QY 356 TTGGTCTTGAAGTGTTCGAGGAAAGCCAGTTTTTGAACTCTCTGGTGGCAACCAACAAAC 415
DB 6246 TCGGTCTGGAACATCGGCTTCATAATTTTCTCGCAACTTCTGCTGGCGAGCAGCAAA 6305
QY 416 GAACTGGTTCGCGCGGTACTGCTCAAAATCCCGAATATCTCTGGCTGATGAACCAA 475
DB 6306 GAGTAGCAGATTGCGGTGCCCATAGCCAAAAATCCAAAAATCTTGCTTTGTGATGAACCAA 6365
QY 476 CCGGAGCTCTAGATTTTAAACAAACAGTGTAGTGTATGAGAAGCATTTGAGACACTTCGCGG 535
DB 6366 CCGGTGCCTTGATTTATCATACAGAAAAACAAATCTTGAAAAATCTTCAAGATATGGCCA 6425
QY 536 ACAAAGGCGCCACCGTGTGTGT 558
DB 6426 GAAAAAGAGGAAAAACAGTTATT 6448

[illegible]

Qy	270	GAATCCCGACGAGCAGTTAAATTCACCTTCAGCTTGGCTGGGAAACACACAAATGGCC	329
Db	2432	CATGCTAAAGCTGACACGCTGTCAAAATGTTGAACCTGCTCTGACCTTACATGAAGTCTC	2491
Qy	330	TGAAAT-----TCCTCAAGTACTTTCATGCTTGTGTTGGTCTTGAATCGTT	371
Db	2492	TAAAAAGAGAGACGCGCGCTTTGGAAATGCTGCGATTGTTGCGTCTGGAAGAGAG	2551
Qy	372	CGAGGAAAAAGCCAGTTTTTGAACCTCTCTGGTGGCGAACAACAACGACTCGCTTGGCCCG	431
Db	2552	AAGCGACTTTAAAGCCTATGAGCTGTGCGCGGCTCAAAAGCAACGGGTAGCTATTGCGCG	2611
Qy	432	GGTACTGCTCAAAAAATCCCGGAATAATCTGGCTGTATGAACCAACCGGAGCTCTAGATT	491
Db	2612	TGCTCTGGTGACCAATCCAGCTTTATTCTGGAGACGAGCGGACTGGAGCGCTGATAC	2671
Qy	492	AACAAACAGTGAAGCTAGTCTATAGAACANTGAGAGCACTCGCGCAACAAAGCGCCACCGT	551
Db	2672	CAAGACAGTGTTCAGATCATGAGTGTTCAGACGAGTTTAACGAAACAAGGCAAACTAT	2731
Qy	552	TGTTGTGCTACGACTCGCC	572
Db	2732	TGTCATCATTCACCAAGCGCC	2752
RESULT 15			
AE006565/c			
LOCUS	AE006565	10738 bp	DNA linear BCT 01-JUN-2001
DEFINITION	Streptococcus pyogenes M1 GAS strain SF370, section 94 of 167 of the complete genome.		
ACCESSION	AE006565 AE004092		
VERSION	AE006565.1 GI:13622379		
KEYWORDS			
ORGANISM	Streptococcus pyogenes M1 GAS		
REFERENCE	Streptococcus pyogenes M1 GAS		
AUTHORS	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.		
	1 (bases 1 to 10738)		
	Ferretti,J.J., McShan,W.M., Adjic,D., Savic,D., Savic,G., Lyon,K., Primeaux,C., Sezate,S.S., Surorov,A.N., Kenton,S., Lai,H., Lin,S., Qian,Y., Jia,H.G., Najjar,F.Z., Ren,Q., Zhu,H., Song,L., White,J., Yuan,X., Clifton,S.W., Roe,B.A. and McLaughlin,R.E.		
TITLE	Complete genome sequence of an M1 strain of Streptococcus pyogenes		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 98 (8), 4658-4663 (2001)		
MEDLINE	21192684		
PUBMED	11296296		
REFERENCE	2 (bases 1 to 10738)		
AUTHORS	Ferretti,J.J., McShan,W.M., Adjic,D., Savic,D., Savic,G., Lyon,K., Primeaux,C., Sezate,S.S., Surorov,A.N., Kenton,S., Lai,H., Lin,S., Qian,Y., Jia,H.G., Najjar,F.Z., Ren,Q., Zhu,H., Song,L., White,J., Yuan,X., Clifton,S.W., Roe,B.A. and McLaughlin,R.E.		
TITLE	Direct Submission		
JOURNAL	Submitted (10-APR-2001) Department of Microbiology and Immunology, University of Oklahoma Health Sciences Center, 940 SL Young Blvd, Oklahoma City, OK 73104, USA		
FEATURES	Location/Qualifiers		
source	1..10738		
	/organism="Streptococcus pyogenes M1 GAS"		
	/mol_type="genomic DNA"		
	/strain="SF370"		
	/serotype="M1"		
	/db_xref="taxon:160490"		
	complement(89..1429)		
gene	/gene="SPY1252"		
	complement(89..1429)		
CDS	/gene="SPY1252"		
	/note="Best Blast hit = sp P47394 Y148_MYCGE_HYPOTHETICAL PROTEIN MG148 - sp 1361569 pir D64216 hypothetical protein homolog MG148 - Mycoplasma genitalium		
	-gi 3844741 gb AAC71366.1 (U39695) conserved hypothetical protein [Mycoplasma genitalium]"		
	/codon_start=1		
	/transl_table=11		


```
gene      complement(8639..8839)
CDS       /gene="SPY1261"
          complement(8639..8839)
          /gene="SPY1261"
          /note="Best Blastp hit = emb|CA885690.1| (AJ271696)
          mismatched base pair and cruciform DNA recognition protein
          (Agaricus bisporus)"
          /codon_start=1

Query Match      13.1%; Score 80; DB 1; Length 10738;
Best Local Similarity 49.0%; Pred. No. 3.7e-13;
Matches 243; Conservative 0; Mismatches 250; Indels 3; Gaps 1;

QY 66  TAAGTTTTTACCAGGAACAACTGACACTGAGCGTCCGGTTCAGGAAATCGAC 125
Db 6450  TGAGATTATAAAGAGAACTAGTGGTTATTTGGGAGCATCCGGAGCTGGTAAATCGAC 6391
QY 126  TTGCTCAACTGCTTGGCACATTGACAAACCAAGTTCGGGACAGATCCTTGTGAGGA 185
Db 6390  AGTCCTCAATATTTTGGGGGCGATGGACACAGTAGATGCAAGTCAAGTAATTATTGATGG 6331
QY 186  TGTAGACCTTCTGAAACTCTCTACGGGTAAAGCAACGGTTATACAGGAAAAATACGGTGGG 245
Db 6330  CAAAGACATTTGCTCAATTACAGGCTTAAGGCTTTAACTCAGTATCGTCGGAATGCTATTGG 6271
QY 246  CTATTTATTTCAAGATTATGCCCTTGATTCGCCGACAGGACAGTTAAATTCACCTTCAGCT 305
Db 6270  TTTCGTTTTTCAATTTTATAAATTTGGTTCTCTAATTTGACAGCTAAAGAAAAATGTTGAAT 6211
QY 306  TCGGTTGGAAAAACACAAATGGCTGAAATTCCT---CAAGTACTTTCATGCTGTGGTCT 362
Db 6210  AGCAGTTGAAATTTGTAGCAGATGCTTTAGATCCTGTGACCATTTTAAAGGAAGTAGGACT 6151
QY 363  TGAGTCGTTTCAGGAAAGCCAGTTTTTTGAACTCTCTGGTGGCGAAACAACAGAACTGC 422
Db 6150  CAGTCATCGTCTGGATCAATTTCTCTGCTCAGCTCTCAGGTGGTGAACAGCAACGGGTTTC 6091
QY 423  GTTGGCCCGGTACTGCTCAAAATCCCGAATAATTCGGCTGATGAACCAACCGGAGC 482
Db 6090  GATAGCACGTGCTTACGTAATAAACCCCTAAATTTGCTTTTGTGATGAACCTACAGTGC 6031
QY 483  TCTAGATTTTAACAAACAGTGAAGTGTATGATAGAGCAATTGAGAGCACTCGCCGACAAAGG 542
Db 6030  CCTTGACTACCAACAGGAAGCAAAATTTCTAACCCCTTTACAGGATATGSCACAAACTAA 5971
QY 543  CGCCACCGTTGTTGTT 558
Db 5970  GGGGACCACGGTAGTT 5955
```

Search completed: July 30, 2004, 13:27:45
Job time : 3638 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 30, 2004, 06:49:19 ; Search time 450 Seconds

(without alignments)
5749.228 Million cell updates/sec

Title: US-09-868-338-7_COPY_1117_1725

Perfect score: 609

Sequence: 1 atgataagaatcaatgacct.....cggatacattatcaacta 609

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124093041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq_29Jan04:*

1: Geneseqn1980s:*

2: Geneseqn1990s:*

3: Geneseqn2000s:*

4: Geneseqn2001as:*

5: Geneseqn2001bs:*

6: Geneseqn2002s:*

7: Geneseqn2003as:*

8: Geneseqn2003bs:*

9: Geneseqn2003cs:*

10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	609	100.0	2370	3 AA65395	Aa65395 Brevibact
2	80.8	13.3	110000	6 ABQ67196_3	Continuation (4 of
3	80.8	13.3	110000	6 ABQ69245_29	Continuation (30 o
4	80.6	13.2	110000	6 ABA90521_11	Continuation (12 o
5	80.4	13.2	1010	2 AAV65222_11	Aav65222 DNA encod
6	80	13.1	711	6 ABN66720	Abn66720 Streptoco
7	79.2	13.0	110000	6 ABA03041_28	Continuation (29 o
8	77.4	12.7	711	4 AAS55747	Aas55747 Streptoco
9	75.6	12.4	750	6 ABN67102	Abn67102 Streptoco
10	75.6	12.4	110000	6 ABN71527_09	Continuation (10 o
11	75.4	12.4	708	7 ABX07412	Abx07412 S. pneumo
12	75.4	12.4	21706	2 AAV52169	Aav52169 Streptoco
13	75.4	12.4	110000	7 ABB56454_14	Continuation (15 o
14	75	12.3	12278	6 ABQ71073	Abq71073 Listeria
15	74.6	12.2	1401	7 ACA39374	Acc39374 Prokaryot
16	74.6	12.2	31241	7 ACC69145	Acc69145 M. genita
17	74.6	12.2	80073	2 AAT58840_5	Continuation (6 of
18	74.6	12.2	110000	6 AAT58840_0	Aat58840 Mycoplasma
19	74.6	12.2	110000	6 ABA03041_12	Continuation (13 o
20	74.4	12.2	756	7 ACA45798	Aca45798 Prokaryot
21	74.4	12.2	110000	6 ABA90521_06	Continuation (7 of
22	73.6	12.1	5312	6 ABQ71007	Abq71007 Listeria
23	73.2	12.0	684	7 ACA42947	Aca42947 Prokaryot

C 24	73.2	12.0	110000	6 ABA90521_17	Continuation (18 o
C 25	73.2	12.0	110000	6 ABA90521_18	Continuation (19 o
26	72.8	12.0	110000	8 ADB12064_06	Continuation (7 of
27	72.6	11.9	672	6 ABN67103	Abn67103 Streptoco
28	72.4	11.9	666	5 AAH65921	Aah65921 C glutami
29	72.4	11.9	789	4 AAF67963	Aaf67963 Corynebact
C 30	72.4	11.9	349980	5 AAH68526	Aah68526 C glutami
C 31	72.4	11.9	349980	5 AAH68527	Aah68527 C glutami
32	71.8	11.8	813	9 ADC93938	Adc93938 E. faeciu
33	71.6	11.8	1062	7 ACA45072	Aca45072 Prokaryot
34	71.2	11.7	708	6 ABN67308	Abn67308 Streptoco
35	71.2	11.7	771	6 ABN92300	Abn92300 Staphyloc
36	71.2	11.7	3319	4 AAH55006	Aah55006 S. epider
C 37	71.2	11.7	3333	4 AAH55024	Aah55024 S. epider
38	71.2	11.7	3657	4 AAH54524	Aah54524 S. epider
39	71.2	11.7	4126	4 AAH54790	Aah54790 S. epider
C 40	71.2	11.7	110000	6 ABQ69245_18	Continuation (19 o
41	71.2	11.7	319630	6 ABQ67194	Abq67194 Listeria
42	70.8	11.6	501	8 ADB08423	Adb08423 Alloioococ
43	70.8	11.6	501	8 ADB08421	Adb08421 Alloioococ
C 44	69.8	11.5	61557	4 AAS59521	Aas59521 Propionib
C 45	69.8	11.5	61557	7 ACF64450	Acf64450 Propionib

ALIGNMENTS

RESULT 1	AAA65395	standard; DNA; 2370 BP.
ID	AAA65395	standard; DNA; 2370 BP.
XX	XX	
AC	AAA65395;	
XX	XX	
DT	15-SEP-2003	(revised)
DT	09-NOV-2000	(first entry)
XX	XX	
DE	Brevibacterium lactofermentum	gltBD gene DNA sequence SEQ ID NO:7.
XX	XX	
XX	Brevibacterium lactofermentum; ABC transporter; breeding; AIPase;	
KW	corynebacterium; gltBD; L-glutamic acid; ds.	
XX	XX	
OS	Corynebacterium glutamicum.	
XX	XX	
FH	Key	Location/Qualifiers
FT	CDS	1..1104
FT	FT	/*tag= a
FT	FT	/product= "ORF1"
FT	FT	1117..1728
FT	FT	/*tag= b
FT	FT	/product= "ORF2"
FT	FT	1759..2370
FT	FT	/*tag= c
FT	FT	/product= "ORF3"
XX	WO200037647-A1.	
ED	29-JUN-2000.	
XX	XX	
PF	16-DEC-1999;	99WO-JP007079.
XX	XX	
PR	18-DEC-1998;	98JP-00360621.
XX	XX	
PA	(AJIN) AJINOMOTO CO INC.	
XX	XX	
PI	Kanno S, Kimura E, Matsui K, Nakamatsu T;	
XX	XX	
DR	WPI; 2000-452189/39.	
DR	P-PSDB; AAB12591, AAB12592, AAB12593.	
XX	XX	
PT	ABC transporter constituent of Brevibacterium lactofermentum, its encoded	
PT	gene and variants, applicable in breeding Corynebacteria particularly for	
XX	production of L-glutamic acid.	
XX	XX	

PS Claim 14; Page 23-28; 34pp; Japanese.

XX The present invention describes a protein (I) which can be used to
CC construct an ATP binding cassette (ABC) transporter. ABC transporter DNAs
CC can be used in breeding Corynebacteria particularly for production of L-
CC glutamic acid. The present sequence encodes three ORFs (open reading
CC frames) from the Brevibacterium lactofermentum gltBD gene, which is used
CC in the exemplification of the present invention. (Updated on 15-SEP-2003
CC to standardise OS field)

SQ Sequence 2370 BP; 633 A; 563 C; 515 G; 659 T; 0 U; 0 Other;

Query Match 100.0%; Score 609; DB 3; Length 2370;
Best Local Similarity 100.0%; Pred. No. 2.2e-191;
Matches 609; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGATAGAAATCAATGACCTCAAGAAATCTTTTCGGCTTCGGATCTTATGGCAAGGTCTC 60
DB 1117 ATGATAGAAATCAATGACCTCAAGAAATCTTTTCGGCTTCGGATCTTATGGCAAGGTCTC 1176

QY 61 AGTCATAAGTTTTACAGGAAACATGACAGCACTGACTGGAGCGTCCGGTTCAGGAAAA 120
DB 1177 AGTCATAAGTTTTACAGGAAACATGACAGCACTGACTGGAGCGTCCGGTTCAGGAAAA 1236

QY 121 TCGACTTTGCTCAACTGCTTGGCAGCACTTGACAAACCAAGTTCGGACAGATCCCTTGTC 180
DB 1237 TCGACTTTGCTCAACTGCTTGGCAGCACTTGACAAACCAAGTTCGGACAGATCCCTTGTC 1296

QY 181 GAGGATGAGACCTTCTCAAACTCTCTACGGGTAAAGCAACGGTTATACAGGAAAAATACG 240
DB 1297 GAGGATGAGACCTTCTCAAACTCTCTACGGGTAAAGCAACGGTTATACAGGAAAAATACG 1356

QY 241 GTGGCTATTTATTCAGATTTAGCTTGNATCCGACAGGACAGTTAAATTCACCTT 300
DB 1357 GTGGCTATTTATTCAGATTTAGCTTGNATCCGACAGGACAGTTAAATTCACCTT 1416

QY 301 CAGCTTGGCGTGGAAAAACAAATGCGCTCAAAATCTCTCAAGTACTTCATGCTGTGGT 360
DB 1417 CAGCTTGGCGTGGAAAAACAAATGCGCTCAAAATCTCTCAAGTACTTCATGCTGTGGT 1476

QY 361 CTGAGTCTGCGAGGAAAGCCAGTTTGTGAATCTCTGCTGGCGGCAACCAACGAACT 420
DB 1477 CTGAGTCTGCGAGGAAAGCCAGTTTGTGAATCTCTGCTGGCGGCAACCAACGAACT 1536

QY 421 CGCTTGGCCCGGTACTGCTCAAAATCCCGGAATAATCTGCGCTGATGAACCAACCGGA 480
DB 1537 CGCTTGGCCCGGTACTGCTCAAAATCCCGGAATAATCTGCGCTGATGAACCAACCGGA 1596

QY 481 GCTCTAGATTTAAACAAAGTGAAGTATGATAGAGCACTTCGCGGCAAAA 540
DB 1597 GCTCTAGATTTAAACAAAGTGAAGTATGATAGAGCACTTCGCGGCAAAA 1656

QY 541 GGCGCCACCGTGTGTGTGCTACCACTCGCCCTCTTCGAGAAATCAGCGGATACCACT 600
DB 1657 GGCGCCACCGTGTGTGTGCTACCACTCGCCCTCTTCGAGAAATCAGCGGATACCACT 1716

QY 601 ATCAAACTA 609

DB 1717 ATCAAACTA 1725

RESULT 2

ABQ67196_3
Continuation (4 of 7) of ABQ67196 from base 300001 (Listeria innocua contig DNA sequence
WP Sequence split into 7 fragments LOCUS ABQ67196 Accession Abq67196

WP Fragment Name Begin End

WP ABQ67196_0 1 110000

WP ABQ67196_1 100001 210000

WP ABQ67196_2 200001 310000

WP ABQ67196_3 300001 410000

WP ABQ67196_4 400001 510000

WP ABQ67196_5 500001 610000

WP ABQ67196_6 600001 694707

Query Match 13.3%; Score 80.8; DB 6; Length 110000;

Best Local Similarity 47.2%; Pred. No. 8.3e-15;

Matches 284; Conservative 0; Mismatches 312; Indels 6; Gaps 1;

QY 1 ATGATAGAAATCAATGACCTCAAGAAATCTTTTCGGCTTCGGAATCTTATGGCAAGGTCTC 60

DB 46005 ATGATGAAATAGTTAATGTCAGTAAAGATACAGATAAATGATTTTGGAAAAAGTC 46064

QY 61 AGTCATAAGTTTTTACCAGGAACATGACAGCACTGACTGGAGCGTCCGGTTCAGGAAAA 120

DB 46065 TCTCTTTCTATTTGGAGCTGGCGAATTTTATTCAGCTCGTCGCGAGAGTGGTAGTGGGAG 46124

QY 121 TCGACTTTGCTCAACTGCTTGGCAGCACTTGACAAACCAAGTTCGGGACAGATCCCTTGTC 180

DB 46125 ACAACGCTGCTAAATGTTTATCGGACACCTAGATTCAAAGATAGTGGGCAAGTTATTATT 46184

QY 181 GAGGATGAGACCTTCTCAAACTCTCTACGGGTAAAGCAACGGTTATACAGGAAAAATACG 240

DB 46185 AACGAGATGGGAATATCAGAAAAAGAGGTATGACTCTTAAAAAAGAGGTG 46238

QY 241 GTGGCTATTTATTTCAAGATTTATSCCTTGATTTCCGACAGGACAGTTAAATTCACCTT 300

DB 46239 TTAGGTTTATTTATTTCAAAATTTATCTATTGATGGAATAAGAAACAGTGTAGAAAACTTA 46298

QY 301 CAGCTTGGCGTGGAAAAACAAATGCGCTCAAAATCTCTCAAGTACTTCATGCTGTGGT 360

DB 46299 TCCATTACAGGTGGGAAAAATCGAAGCTGATGATAGACATTTTGGAGAAAGTAGGAATG 46358

QY 361 CTGAGTCTGCTGAGGAAAAAGCCAGTTTTCGAACTCTCTGCTGGCGGCAACCAACGAACT 420

DB 46359 GATGAGAGCTATTATAGCAAAAAAGATATACCAATTAAGTGTGGAGAAAAACACGAT 46418

QY 421 GCGTGGCCCGGTACTGCTCAAAATCCCGGAATAATCTGCTGATGAACCAACCGGA 480

DB 46419 GCTATTGTGCGCATTTTACTCAAACTTTCACTTTGCGGACGAAACCAACTGGC 46478

QY 481 GCTCTAGATTTAAACAAAGTGAAGTATGATAGAGCACTTCGCGGCAAAA 540

DB 46479 AATTTAGATGATAAAAACAAACAAAAATCATTGAATTTTCTAGCTTGAAGAACAA 46538

QY 541 GGCGCCACCGTGTGTGTGCTACCACTCGCCCTCTTCGAGAAATCAGCGGATACCACT 600

DB 46539 GCTAAGACTTCTGTTGTGTGTCACGATGACCCGGAATATCAGGAAAAAGCATCGGTC 46598

QY 601 AT 602

DB 46599 AT 46600

RESULT 3

ABQ69245_29/c
Continuation (30 of 31) of ABQ69245 from base 2900001 (Listeria innocua DNA sequence #68;
WP Sequence split into 31 fragments LOCUS ABQ69245 Accession Abq69245

WP Fragment Name Begin End

WP ABQ69245_0 1 110000

WP ABQ69245_01 100001 210000

WP ABQ69245_02 200001 310000

WP ABQ69245_03 300001 410000

WP ABQ69245_04 400001 510000

WP ABQ69245_05 500001 610000

WP ABQ69245_06 600001 710000

WP ABQ69245_07 700001 810000

WP ABQ69245_08 800001 910000

WP ABQ69245_09 900001 1010000

WP ABQ69245_10 1000001 1110000

WP ABQ69245_11 1100001 1210000

WP ABQ69245_12 1200001 1310000

WP ABQ69245_13 1300001 1410000

WP ABQ69245_14 1400001 1510000

WP ABQ69245_15 1500001 1610000

WP ABQ69245_16 1600001 1710000

WP ABQ69245_17 1700001 1810000

```
WP ABQ69245_18 1800001 1910000
WP ABQ69245_19 1900001 2010000
WP ABQ69245_20 2000001 2110000
WP ABQ69245_21 2100001 2210000
WP ABQ69245_22 2200001 2310000
WP ABQ69245_23 2300001 2410000
WP ABQ69245_24 2400001 2510000
WP ABQ69245_25 2500001 2610000
WP ABQ69245_26 2600001 2710000
WP ABQ69245_27 2700001 2810000
WP ABQ69245_28 2800001 2910000
WP ABQ69245_29 2900001 3010000
WP ABQ69245_30 3000001 311208

Query Match 13.3%; Score 80.8; DB 6; Length 110000;
Best Local Similarity 47.2%; Pred. No. 8.3e-15;
Matches 284; Conservative 0; Mismatches 312; Indels 6; Gaps 1;

QY 1 ATGATAGAAATCAATGACCTCAAGAAATCTTTGGCGTTCCGATCTTATGGCAAGTCTC 60
Db 34809 ATGATGATTAATGATTAATGCTAGTAAAGATACAGATAAATGATTTGGAAAAGTC 34750
QY 61 AGTCATAAGTTTTTACCAGAACATGACAGACATGACTGGAGCGTCCGGTTTCAGAAAA 120
Db 34749 TCTCTTTCTATTGGAGCTGCGCAATTTATTGAGTCTGTCGCGAGAGTGGTAGTGGAG 34690
QY 121 TCGACTTTGCTCAACTGCTTGGCACATTTGACAAACCAAGTTCCGGACAGATCCTTGTC 180
Db 34689 ACACCTGCTTAATGTTATCGGACACCTAGATTCAAAGATAGTGGCAAGTTATTAT 34630
QY 181 GAGGATGTAGACCTTCTGAAACTCTTACCGGTAAAGCAACGGTTTACAGGAAAAATACG 240
Db 34629 AACGAGATGGAATATCAGA-----CGAAAAAGAGGTTATGACTCTAAAAAAGAGGTG 34576
QY 241 GTGGGCTATTATTTCAGANTATGCTTGTATCCGACAGACAGATTAAATTCACCTT 300
Db 34575 TTAGGTTTATATTTCANAAATTTATCTATTGATGGAATGAACAGTGTGAAAACCTTA 34516
QY 301 CAGCTTGGCGTGGAAAAACAAATCGCCTGAAATTCCTCAAGTACTTCAATGCTGTGGT 360
Db 34515 TCCATTACAGTGGGGAATTCGACGCTGATGATAGCAATTTGGAAGAGTAGGAATG 34456
QY 361 CTGAGCTGTTGAGAAAAAGCGATTTTGAATCTCTGGTGGGGAACAAACGAACCT 420
Db 34455 GATGAGAGCTATTTAGCAAAAAAGTATACCAATTAAGTGGTGGAGAAAAACACGGATT 34396
QY 421 GCGTTGCGCGGTACTGCTCAAAAATCCCGAATAATCTGCTGATCAACCAACCGGA 480
Db 34395 GCTATTGTGGCAATTTACTCAACCACTTCAACTTTTACTTGGGACGAACCACTGGC 34336
QY 481 GCTCTAGATTAAACAAACAGTAGCTAGTCATAGAGCAATTGAGACACTCCGCGACAAA 540
Db 34335 AATTAGATGATAAAACAAACAAAAATCAITGAAATTTATTTCTAGCCCTTGAAGAACAA 34276
QY 541 GCGGCCACCGTTGTTGTTGCTACGCACTCGCCCTCTCCGAGATCAGCGGATACCAT 600
Db 34275 GGTAAAGACTATGTTTGTGTGTCAGCATGACCCGGAATATCAGGAAAAAGCAGATCGGTC 34216
QY 601 AT 602
Db 34215 AT 34214
```

```
RESULT 4
ABA90521_11
Continuation (12 of 24) of ABA90521 from base 1100001 (Genomic sequence of Lactococcus
WP Sequence split into 24 fragments LOCUS ABA90521 Accession ABA90521
WP Fragment Name Begin End
WP ABA90521_00 1 110000
WP ABA90521_01 100001 210000
WP ABA90521_02 200001 310000
WP ABA90521_03 300001 410000
WP ABA90521_04 400001 510000
```

```
WP ABA90521_05 500001 610000
WP ABA90521_06 600001 710000
WP ABA90521_07 700001 810000
WP ABA90521_08 800001 910000
WP ABA90521_09 900001 1010000
WP ABA90521_10 1000001 1110000
WP ABA90521_11 1100001 1210000
WP ABA90521_12 1200001 1310000
WP ABA90521_13 1300001 1400001
WP ABA90521_14 1400001 1510000
WP ABA90521_15 1500001 1610000
WP ABA90521_16 1600001 1710000
WP ABA90521_17 1700001 1810000
WP ABA90521_18 1800001 1910000
WP ABA90521_19 1900001 2010000
WP ABA90521_20 2000001 2110000
WP ABA90521_21 2100001 2210000
WP ABA90521_22 2200001 2310000
WP ABA90521_23 2300001 236589

Query Match 13.2%; Score 80.6; DB 6; Length 110000;
Best Local Similarity 48.9%; Pred. No. 9.6e-15;
Matches 246; Conservative 0; Mismatches 254; Indels 3; Gaps 1;

QY 59 TCAGTCATAAGTTTTTTACCAGGAACAATGACAGCACTGACTGGAGCCTCCGGTTTCAGAA 118
Db 10785 TAAATTTTGAATTTGAATAAGGAGATTTGGCTGTCTCATTTTAGGTTCTCTGGGGCAGGTA 10844
QY 119 ANTGCRCTTTTCTCAACTGCTTTGGCACACTTGCACAAACCAAGTTCGGGACAGATCCTTG 178
Db 10845 AATCAACAGCTCTAAATATTTTAGGAGGAATGGATCAAAATGATGAGGGAGAGGTTATTA 10904
QY 179 TCGAGGATGTAGACTCTTCTGAACTCTCTACGGGTAAAGCAACCGTTATACAGGAAAAATA 238
Db 10905 TTGATGAGGAAAAAATCTCTAAATTTTCAAACAAAGAAATGATTACCTACCGAGTTATG 10964
QY 239 CGGTGGCGCTATTATTTCAAGATTTATGCTTGTATCCGACAGACAGATTAATTTCAACC 298
Db 10965 CCGTTGGTTTGGTTTTTCAATTTTATAATCTGTCAATAATTTGACAGCACTAGANAATG 11024
QY 299 TTCAGCTTGGGTGAAAAACACAAATGGCTGAAATTCCTC---AAGTACTTTCATGCTG 355
Db 11025 TCGAATTAGCTTCTGAAATTTGTCGAAATGCTCTTGATGCTAAAGAAGTCTGAGAGTG 11084
QY 356 TTGCTCTTGAGTCGTTTCGAGGAAAGCCAGTTTGTGAATCTCTCTGTTGGGGAACAAAC 415
Db 11085 TCGGTCTGGAACATCGGCTTCATAATTTTCTCTGCAACTTCTGTGTGGGAGCAGCAAA 11144
QY 416 GAACTGCTTGGCCCGGCTACTGCTCAAAAATCCCGAATAATTTCTGCTGATGAACCAA 475
Db 11145 GAGTAGCGATTGCGCGTCCCATGCCCAAAATCCAAAATCTTGTCTTTGATGAACCAA 11204
QY 476 CCGGAGCTCTAGATTTAAACAAACAGTGAAGTCTATAGAGCAATTGAGAGCACTCGCCG 535
Db 11205 CCGTGCCTTGGATTATCATACAGGAAAAACAATCTTGAANAATCTTCAAGATATGSCCA 11264
QY 536 ACAAGGCGCCACCGTTGTTGTT 558
Db 11265 GAAAAGAGGAAAAACAGTTATT 11287
```

```
RESULT 5
AAV65222/c
ID AAV65222 standard; DNA; 1010 BP.
XX
AC AAV65222;
XX
DT 24-DEC-1998 (first entry)
XX
DE DNA encoding a S. pneumoniae transport protein.
XX
KW Streptococcus pneumoniae protein; recombinant; gene expression; DNA chip;
virulence; antibody; infection; detection; treatment; ss.
```

XX OS Streptococcus pneumoniae.
 XX FN WO9826072-A1.
 XX PD 18-JUN-1998.
 XX PF 09-DEC-1997; 97WO-US022578.
 XX PR 13-DEC-1996; 96US-0036281P.
 XX PA (ELIL) LILLY & CO ELI.
 XX PI Baltz RH, Burgett SG, Dehoff BS, Hoskins JA, Jaskunas SR;
 XX PI Mills BJ, Norris FH, Peery RB, Rostek PR, Skatrud PL;
 XX PI Smith MC, Solenberg PJ, Treadway PJ, Young Bellido ML;
 XX DR WPI: 1998-348529/30.
 XX DR P-PSDB; AAW80648.
 XX ST Streptococcus pneumoniae nucleic acid sequences - used in DNA chips for
 PT evaluating gene expression, and identification of virulence genes.
 XX PF
 XX PS Claim 1; Page 63; 333pp; English.
 XX CC This DNA sequence encodes a Streptococcus pneumoniae transport protein.
 CC The invention provides DNA sequences (AAV65201 to AAV65304) from the
 CC Streptococcus pneumoniae genome and corresponding protein sequences
 CC (AAW80605 to AAW80728). A recombinant host containing a vector comprising
 CC any of the above nucleic acids can be used for the recombinant expression
 CC of the protein sequences. The invention also provides a DNA chip having
 CC arrayed on it at least 15 base pair fragment of any one or more of these
 CC DNA sequences. The DNA chip can be used for evaluating gene
 CC expression in S. pneumoniae and for identifying virulence genes in S.
 CC pneumoniae. Antibodies that selectively bind to the above proteins or
 CC peptide fragments can be used to treat S. pneumoniae infection. The
 CC antibodies can also be used to detect S. pneumoniae cells
 XX SQ
 XX Sequence 1010 BP; 285 A; 247 C; 207 G; 271 T; 0 U; 0 Other;
 Query March 13.2%; Score 80.4; DB 2; Length 1010;
 Best Local Similarity 49.8%; Pred. No. 9.9e-16;
 Matches 231; Conservative 0; Mismatches 231; Indels 2; Gaps 1;
 QY 55 GGTCTCAGTCATAGTTTATACAGGAAACCAATGACAGCACTGACTGAGCGTCCGGTTCA 114
 DB 1007 GATGTGAATTTGAGATTGAAGGGGGAAGTGGTTATTATCTTGTGCTTCAGGTGCA 948
 QY 115 GGAAATCGAGTTTGTCTCACTGCTTGGCACACTTGACAAACAGATTCGCGACAGATC 174
 DB 947 GGCAAGTCAACAGTCTTAACCTTCTTGGGGGAATGGATACCAATGATGAAGGGGAATC 888
 QY 175 CTTCTCAGGATGATGAGACCTCTGAAACTCTCTACCGGTAAAGCAACGGTTATACAGCAA 234
 DB 887 TGGATTGATGGTGTATATATGCGGATATAGTTCACAGCGCCCAATACCTAGA 828
 QY 235 AATACGTTGGGCTATTTATTTCAAGATATGCTTGAATCCGACAGGACAGTTAAATTC 294
 DB 827 AATGATGTGGGTTTGTGTTTTCAGTTTATTAATCTAGTTTCTAATCTGACAGTAAGGAA 768
 QY 295 AACCTTCAGTTCGGGGGAAACACAAATGCCTGGAATTC-CTCAAGTACTTCATG 352
 DB 767 AATGTGAACGCTCTGAAATTTGTGACATGCTTGAAATCTGATCAGGCTTTCACAG 708
 QY 353 CTGTGCTTGTAGTCTGTTCCAGAAAGCCAGTTTTTGAATCTCTGTGGCGGAACAAC 412
 DB 707 ATGTAGTCTGCTCATCGCTAAATAAATTTCCAGCCCAAGCTTCTGAGGGGAGCAAC 648
 QY 413 AACCAACTGGTGGCCCGGCTACTGCTCAAAATCCCGAATAATCTGGCTGATGAAC 472
 DB 647 AGCGAGTCTCATTTGACGCGCGGTGACCCAAATCTTAAATCTCTTCTTGTGATGAAC 588
 QY 473 CAACCGGAGCTCTAGATTTTAAACAAAGTGTAGTGTAGTATAGAA 516

DB 587 CGACTGGAGCCTTGGATTATCAGACGGCAAGCAGGTTTGGAA 544
 RESULT 6
 ID AEN66720 standard; DNA; 711 BP.
 XX AC AEN66720;
 XX DT 01-JUL-2002 (first entry)
 XX DE Streptococcus polynucleotide SEQ ID NO 1353.
 XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
 XX OS Streptococcus pyogenes.
 XX PN WO200234771-A2.
 XX PD 02-MAY-2002.
 XX PF 29-OCT-2001; 2001WO-GB004789.
 XX PR 27-OCT-2000; 2000GB-00026333.
 XX PR 24-NOV-2000; 2000GB-00028727.
 XX PR 07-MAR-2001; 2001GB-00005640.
 XX PA (CHIR-) CHIRON SPA.
 XX PA (GENO-) INST GENOMIC RES.
 XX PI Telford J, Masignani V, Margarit Y RosI, Grandi G, Fraser C;
 PI Tettelin H;
 XX WPI: 2002-352536/38.
 XX P-PSDB; ABP26089.
 XX PT New Streptococcus protein for the treatment or prevention of infection or
 PT disease caused by Streptococcus bacteria, such as meningitis, and for
 PT detecting a compound that binds to the protein.
 XX PS Claim 7; Page 3291; 4525pp; English.
 XX CC The invention relates to a protein (ABP25413-ABP30895) from group B
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 CC the specification. The proteins have antibacterial and antiinflammatory
 CC activity. (I), nucleic acids encoding (I), AEN66044-ABN71546 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins
 XX SQ
 XX Sequence 711 BP; 211 A; 130 C; 159 G; 211 T; 0 U; 0 Other;
 Query Match 13.1%; Score 80; DB 6; Length 711;
 Best Local Similarity 49.0%; Pred. No. 1.1e-15;
 Matches 243; Conservative 0; Mismatches 250; Indels 3; Gaps 1;
 QY 66 TAAGTTTTTACAGGAAACATGACAGCACTGAGCGTCCGGTTCAGGAAATCGAC 125
 DB 96 TGAGATTAAAGAGAACTAGTGGTTATTTGGGAGCATCCGAGCTGGTAAATCGAC 155

QY 126 TTGCTCACTGCTTGGCAGACTTCACAAACCAAGTTCGGACAGATCCTTGTGAGGA 185
Db 156 AGTCCTCAATATTTTGGGGGCAAGGACACAGTAGAGTCAAGTCAATATTGATGG 215
QY 186 TGTAGACCTTCTGAAACTCTCTACGCGTAAGCAACGTTTATACAGGAAAAATACGGTGG 245
Db 216 CAAAGACATTCCTCAATACAGCTTAAGGCTTAACTCAGTATCGTTCGGAATGCTATTGG 275
QY 246 CTATTTATTTTCAGAGATATGCTTGTGATTCCTCGACAGGACAGTTAAATTTCAACTTCAGCT 305
Db 276 TTTTCGTTTTTCAATTTTATTAATTTGTTCTTAATTTGACAGCTTAAAGAAATGTTGAAT 335
QY 306 TGGGTGGGAAAAACACAAATGGCTGAAATTCCT---CAAGTACTTCACTGCTGTTGGTCT 362
Db 336 AGCAGTTGAAATTTGACAGATCTTTAGATCTCTGTGACCAATTTTAAAGGAAGTAGGACT 395
QY 363 TGAAGTCTGTCAGGAAAGCCAGTTTGTAACTCTCTGTGGGGAACACACAGCACTGC 422
Db 396 CAGTCATCGCTGAGTCAATTTCTCTCTCAGCTCTCAGGTGGTGAACAGCAACGGGTTTC 455
QY 423 GTTGGCCCGGGTACTGCTCAAAATCCCCGAATAATTCCTGGCTGATGAACCAACCGGAGC 482
Db 456 GATAGCAGTGCCTTAGCTAAACCCCTAAATTTGCTTCTTTGTGATGATCACTACAGTGC 515
QY 483 TCTAGATTTTAAACAGTGTAGTCTATAGAGCAATTTAGAGCACTCGCCGACAAAGG 542
Db 516 CTTGTACTACCAACAGGAAGCAAAATCTTAACCTCTTACAGGATATGGCACAACATAA 575
QY 543 CGCCACCGTTGTGTT 558
Db 576 GGGGACACCGTAGTT 591

RESULT 7
ABA03041_28/c
Continuation (29 of 30) of ABA03041 from base 2800001 (Listeria monocytogenes BGD-e gene)
WP Sequence split into 30 fragments LOCUS ABA03041 Accession ABA03041

WP	Fragment Name	Begin	End
WP	ABA03041_00	1	110000
WP	ABA03041_01	100001	210000
WP	ABA03041_02	200001	310000
WP	ABA03041_03	300001	410000
WP	ABA03041_04	400001	510000
WP	ABA03041_05	500001	610000
WP	ABA03041_06	600001	710000
WP	ABA03041_07	700001	810000
WP	ABA03041_08	800001	910000
WP	ABA03041_09	900001	1010000
WP	ABA03041_10	1000001	1110000
WP	ABA03041_11	1100001	1210000
WP	ABA03041_12	1200001	1310000
WP	ABA03041_13	1300001	1410000
WP	ABA03041_14	1400001	1510000
WP	ABA03041_15	1500001	1610000
WP	ABA03041_16	1600001	1710000
WP	ABA03041_17	1700001	1810000
WP	ABA03041_18	1800001	1910000
WP	ABA03041_19	1900001	2010000
WP	ABA03041_20	2000001	2110000
WP	ABA03041_21	2100001	2210000
WP	ABA03041_22	2200001	2310000
WP	ABA03041_23	2300001	2410000
WP	ABA03041_24	2400001	2510000
WP	ABA03041_25	2500001	2610000
WP	ABA03041_26	2600001	2710000
WP	ABA03041_27	2700001	2810000
WP	ABA03041_28	2800001	2910000
WP	ABA03041_29	2900001	2944528

Query Match 13.0%; Score 79.2; DB 6; Length 110000;
Best Local Similarity 47.0%; Pred. No. 2.8e-14;
Matches 283; Conservative 0; Mismatches 313; Indels 6; Gaps 1;

QY 1 ATGATAGAAATCAATGACCTCAAGAAATCTTTTGGCGTTCGATCTTTATGCGAAGTCTC 60
Db 58203 ATGATTTGAAATTTAGTTTAAATGCTAGTAAAGATACAGATAAATTTGATTTTGGAAAAAGTC 58144
QY 61 AGTCATAGTTTTTACCAGGAACAATGACAGCACTGACTGAGCGTCCGGTTCAGGAAAA 120
Db 58143 TCTCTTTCTATTGGAGCTGGTGAATTTATTGCACTCGTCGGCGAGAGTGGTAGTGGGAAG 58084
QY 121 TCGACTTTGCTCACTGCTTGGCAGCACTTGACAAACCAAGTTCGGGACAGATCCTTGTCT 180
Db 58083 ACCAGCTGTCTAAATGTTATCGGACACCTAGATTCAAAAGATAGTGGGCAAGTTATTATT 58024
QY 181 GAGGATGTAGACCTTCTGAAACTCTCTACGCGTAAGCAACCGTGTATACAGGAAAAATACG 240
Db 58023 AACGAGATGGAATATCAG-----CGAAAAAGAGGTTATGACTCTTAAAAAAGAGGTG 57970
QY 241 GTGGGCTATTTTATTTCAAGATTATGCCCTTGAATCCCGACAGGACAGTTAAATTTCAACCTT 300
Db 57969 TTAGGTTTATTTTCCAAAATTTATCTATTGATGGAATAATGAACAGTGTGTAGAAAACTTA 57910
QY 301 CAGCTTGGCGTGGAAAAACACAAATGGCTGAAATTCCTCAAGTACTTCTCATGCTGTTGGT 360
Db 57909 TCCATTACAGTGGGGAATAATCGCAAGCTGATGATAGAGCATTTAGAAGTAGGATG 57850
QY 361 CTTGAGTGTTCGAGGAAAAAGCCAGTGTGTAACCTCTCTGGTGGCGAACAAACAGAACT 420
Db 57849 GATGAGAGCTATTTAGCAAAAAAAGTATACCAATTAAGCGGTGAGAAAAAACAACGGATT 57790
QY 421 GCGTTGGCCCGGTACTGCTCAAAAAATCCCGAATAATTTCTGGCTGATGAACCAACCGGA 480
Db 57789 GCTATTGCGGCATTTTACTCAAAACCATTTCACTGTTTACTTGGGACGACCAACAGGC 57730
QY 481 GCTCTAGATTTAACAACAGTGTAGTCTATAGAGCATTTGAGAGCACTGCGCGACAAA 540
Db 57729 AACTTGGATGATAAAACAAACAAAAAATCATTTGAATTTATTTTAGCCTTGAAGAAGCAA 57670
QY 541 GCGCCACCGTTGTTGTTGCTACGCACTCGCCCTCTTCCGAGATCAGCGGATACCAT 600
Db 57669 GGTAAAGTATCTGTTGTTGTTTACGATGACCCAGAAATATCTCAAAAGCAGATCGAATC 57610
QY 601 AT 602
Db 57609 AT 57608

RESULT 8
AAS55747
ID AAS55747 standard; DNA; 711 BP.
XX AAS55747;
AC AAS55747;
XX AAS55747;
DT 13-FEB-2002 (first entry)
XX Streptococcus pneumoniae DNA for cellular proliferation protein #318.
DE Antisense; ds; prokaryotic cellular proliferation gene; antibiotic;
DE antibacterial; drug design.
XX Streptococcus pneumoniae.
OS WO200170955-A2.
PN 27-SEP-2001.
PD 21-MAR-2001; 2001WO-US009180.
PF 21-MAR-2000; 2000US-0191078P.
PR 23-MAY-2000; 2000US-0206848P.
PR 26-MAY-2000; 2000US-0207727P.
PR 23-OCT-2000; 2000US-0242578P.
PR 27-NOV-2000; 2000US-0253629P.
PR 22-DEC-2000; 2000US-0257931P.
PR 16-FEB-2001; 2001US-0269308P.

XX PA (ELIT-) ELITRA PHARM INC.
 XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GU;
 XX PI Yamamoto RT, Xu HH;
 XX DR WPI: 2001-611495/70.
 XX DR P-PSDB; AAU37888.
 XX PT New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids.
 XX PS Claim 27; SEQ ID NO 9384; Slipp; English.
 XX The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the genes,
 CC their use in the discovery of novel antibiotics, the essential genes
 CC themselves and the encoded proteins. The prokaryotes used are *Escherichia*
 CC coli, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*,
 CC *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also
 CC useful for the identification of potential new targets for antibiotic
 CC development. The antisense nucleic acids can also be used to identify
 CC proteins used in proliferation, to express these proteins, and to obtain
 CC antibodies capable of binding to the expressed proteins. The proteins can
 CC be used to screen compounds in rational drug discovery programmes. The
 CC antisense nucleic acid sequence is also useful to screen for homologous
 CC nucleic acids which are required for cell proliferation in a wide variety
 CC of organisms. The present sequence encodes an essential prokaryotic
 CC cellular proliferation protein. Note: The sequence data for this patent
 CC did not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX SQ Sequence 711 BP; 187 A; 132 C; 190 G; 202 T; 0 U; 0 Other;
 Query Match 12.7%; Score 77.4; DB 4; Length 711;
 Best Local Similarity 50.8%; Pred. No. 8.2e-15;
 Matches 236; Conservative 0; Mismatches 226; Indels 3; Gaps 2;
 QY 55 GGCTCAGTCATAAGTTTACACAGGAACAATGACAGCACTGACTGGAGCGTCCGGTTCA 114
 Db 82 GATGTGAATTTGAGATTGAAGAGGGGAACTGGTTATTATCTCTTGGTCTTCAGGTGCA 141
 QY 115 GGAAATCGACTTGTCTCACTGCTTGGCACACTTGACAAACCAAGTTCGGACAGATC 174
 Db 142 GCGAAGTCAACAGTTCTTAACCTTCTTGGGGGATGGATACCAATGATGAAGGGGAATC 201
 QY 175 CTGTGCGAGGATGAGACCTTCTGAAACTCTCTACGGGTAAAGCAACGGTTATACAGGAA 234
 Db 202 TGGATTGATGGTGTATATATGCGGATATAGTTCCACAGCGCAACCAATTACCGTAGA 261
 QY 235 AATACGCTGGCTATTTATTCAGATATGCTTGGATTCCTGATCCCGACAGCAGTAATTC 294
 Db 262 AATGATGFGGGTGTGTTTTTCAGTTTATACTAGTTCTTATTCAGCAGTACAGTAAGAA 321
 QY 295 AACCTTCAGCTTGGCGTGGAAA--AACACAAATGGCCTGAAAT--TCCTCAAGTACTTCAT 351
 Db 322 AATGTGAACCTGGCTCTGAAATTTGTGACAGATGCCCTTGAATCTGATCAGGCTTGACA 381
 QY 352 GCTGTGTGCTTGTAGTGTTCGAGGAAAGCCAGTCTTTGAACTCTCTGTGGCGAACRA 411
 Db 382 GATGTAGTCTGGTCTATCGTCTAAATAACTTTCCAGCCAGCTTTCTGAGGGGAGCAA 441
 QY 412 CAACGAACCTGGCTGGCCGGGTACTGCTCAAAAATCCCGAATAATTCCTGGCTGATGA 471
 Db 442 CAGCGAGTCTCCATGTGACGCGCGGTAGCCAAATCTTAATATTCCTCTTTGTGATGA 501
 QY 472 CCACCGGAGCTCTAGATTTTAAACAAACAGTAGGCTAGTATAGAA 516
 Db 502 CCGACTGGAGCTTGGATTATCAGACGGGCAAGCAGGTTTGGAA 546

ABN67102
 ID ABN67102 standard; DNA; 750 BP.
 XX AC ABN67102;
 XX DT 01-JUL-2002 (first entry)
 XX DE Streptococcus polynucleotide SEQ ID NO 2117.
 XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
 KW antinflammatory; infection; vaccine; meningitis; gene therapy; ds.
 XX OS Streptococcus agalactiae.
 XX WO200234771-A2.
 XX 02-MAY-2002.
 XX 29-OCT-2001; 2001WO-GB004789.
 XX 27-OCT-2000; 2000GB-00026333.
 PR 24-NOV-2000; 2000GB-00028727.
 PR 07-MAR-2001; 2001GB-00005640.
 XX (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX Telford J, Masignani V, Margarit Y Rosl, Grandi G, Fraser C;
 PI Tettelin H;
 XX WPI: 2002-352536/38.
 DR P-PSDB; ABP26471.
 XX New Streptococcus protein for the treatment or prevention of infection or
 PT disease caused by Streptococcus bacteria, such as meningitis, and for
 PT detecting a compound that binds to the protein.
 XX Claim 7; Page 3364; 4525pp; English.
 XX The invention relates to a protein (ABP25413-BP30895) from group B
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
 CC the specification. The proteins have antibacterial and antinflammatory
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly *S. agalactiae* and *S. pyogenes*.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins
 XX SQ Sequence 750 BP; 249 A; 119 C; 143 G; 239 T; 0 U; 0 Other;
 Query Match 12.4%; Score 75.6; DB 6; Length 750;
 Best Local Similarity 48.3%; Pred. No. 3.3e-14;
 Matches 259; Conservative 0; Mismatches 259; Indels 18; Gaps 1;
 QY 52 CAAGTCTCAGTCATAAGTCTTTTACAGGAACAATGACAGCACTGAGTCGAGCTCCGGT 111
 Db 73 CAAGAGCTCGATTTTAAAGTTGAACAGGGGAATTCATTGCTATTATGAGAGATCTGGT 132
 QY 112 TCAGGAATTCGACTTTCTGCTCACTGCTTGGCACACTTGACAAACCAAGTTCGGGACAG 171
 Db 133 TCGGAAAAACAACACTCTCTGCTAAATATTTCCTACTTTTAAAAAACCCGCAATGGACAA 192
 QY 172 ATCTCTGTGCGAGGATGTAGACCTTCTGAACTCTCTACGCGTAAAGCAACGGTTATACAGG 231

Db 193 GTGATTTAAATGGGAGATATACGAAATTAAGAGGCAAAATAGCGATTTCT 252
Qy 232 AAAAAATACGGTGGCTATTTATTTCAAGATATGCTTATCCCGACAGCAGTAA 291
Db 253 TTGAAAAATCTTGTTTGGTCTTTTCAAGATTTTCAAGATTTTCAAGATTTT 312
Qy 292 TTCAACCTTCAGCTTCGGTGG-----AAAAACACAAATGGCTGAA 333
Db 313 GATTAATATCTATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 372
Qy 334 ATTCCTCAAGTACTTCATGCTGTGTTGCTTCTGAGTCTGCTCAAAATCCCGA 453
Db 373 TTGTCAGATTTATCTTCTATCTGAGATTTGATGACTTATTAGATAAGAGAC 432
Qy 394 CTCTCTGGTGGCAACAAAGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 453
Db 433 CTCTCTGGTGGCAACAAAGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 492
Qy 454 ATAAATCTGCTGATGAACCAACCGGAGCTCTAGATTTAAACAAAGTGTAGT 513
Db 493 ATTTTATTAGCAGATGAACCAACCGGAGCTTACCGTAAATTCAGAGACT 552
Qy 514 GAAGCATTTAGAGACTTCGCGGCAAAAGCGGCCACCGTTGTTGCTTACGACT 569
Db 553 AATTTATTGAAACTATTACTTTGGATGGACAAACTATTTTGATGTAACCCAT 608

RESULT 10
ABN71527_09
Continuation (10 of 22) of ABN71527 from base 900001 (Streptococcus polynucleotide SEQ 1)
WP Sequence split into 22 fragments LOCUS ABN71527 Accession ABN71527
Fragment Name Begin End
WP ABN71527_01 1 110000
WP ABN71527_02 100001 210000
WP ABN71527_03 200001 310000
WP ABN71527_04 300001 410000
WP ABN71527_05 400001 510000
WP ABN71527_06 500001 610000
WP ABN71527_07 600001 710000
WP ABN71527_08 700001 810000
WP ABN71527_09 800001 910000
WP ABN71527_10 900001 1010000
WP ABN71527_11 100001 1110000
WP ABN71527_12 1100001 1210000
WP ABN71527_13 1200001 1310000
WP ABN71527_14 1300001 1410000
WP ABN71527_15 1400001 1510000
WP ABN71527_16 1500001 1610000
WP ABN71527_17 1600001 1710000
WP ABN71527_18 1700001 1810000
WP ABN71527_19 1800001 1910000
WP ABN71527_20 1900001 2010000
WP ABN71527_21 2000001 2110000
WP ABN71527_22 2100001 2155561

Query Match 12.4%; Score 75.6; DB 6; Length 110000;
Best Local Similarity 45.3%; Pred. No. 4.4e-13;
Matches 259; Conservative 0; Mismatches 259; Indels 18; Gaps 1;

Qy 52 CAAGGCTCAGTCATAGTTTATACAGGAAACATGACAGCACTGACTGAGCGTCCGGT 111
Db 51728 CAAGACGTCGATTTAAAGTTGAACAGAGGGAATTCATGCTATTATGGAGAGTCTGGT 51787
Qy 112 TCAGGAAATCGACTTTGCTCACTGCTTGGGACACTTGACAAACAGTTCGGACAG 171
Db 51788 TCGGAAAAACAACTCTGCTAAATATTTAGTACTTTAGAAAAACCGCAATATGGACAA 51847
Qy 172 ATCTCTTCGAGGATGTAGACCTTCTCAAACTCTACGCGGTAAAGCAACGTTATACAG 231
Db 51848 GTGATTTTAAATGGGAGATATATACGAAATTAAGAGGCAAAATAGGAGTTTCTG 51907
Qy 232 AAAAAATACGGTGGCTATTTATTTCAAGATTTATGCTTGTATCCCGACAGCAGTAA 291

Db 51908 TTGAAAAATCTTGGTTTCTGCTTTTCAGGACTTCAACCTTTTGGATCTCTCTCTAAGA 51967
Qy 292 TTCAACCTTCAGCTTCGGTGG-----AAAAACACAAATGGCCTGAA 333
Db 51968 GATAATATCTATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 52027
Qy 334 ATTCCTCAAGTACTTCATGCTGTGTTGCTTCTGAGTCTGCTCAAAATCCCGA 393
Db 52028 TTGTCAGATTTATCTTCTCACTGAGATTTGATGACTTATTAGATAAGAGACTT 52087
Qy 394 CTCTCTGGTGGCAACAAAGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 453
Db 52088 CTCTCTGGTGGCAACAAAGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 52147
Qy 454 ATAAATCTGCTGATGAACCAACCGGAGCTCTAGATTTAAACAAAGTGTAGT 513
Db 52148 ATTTTATTAGCAGATGAACCAACCGGAGCTTACCGTAAATTCAGAGACTT 52207
Qy 514 GAAGCATTTAGAGCACTTCGCGGCAAAAGCGGCCACCGTTGTTGCTTACGACT 569
Db 52208 AATTTATTGAACTATTACTTTGGATGGACAAACTATTTTGATGTAACCCAT 52263

RESULT 11
ABX07412
ID ABX07412 standard; DNA; 708 BP.
XX
AC ABX07412;
XX
DT 27-OCT-2003 (revised)
DT 11-FEB-2003 (first entry)
XX
DE S. pneumoniae type 4 strain coding region #1700.
XX
KW Gene; ds; bacterial meningitis; pneumonia; sepsis; otitis media;
KW ear infection; antiinflammatory; antibacterial; immunostimulant;
XX
KW auditory; respiratory; gene therapy; vaccine.
OS Streptococcus pneumoniae; type 4 strain.
PN WO200277021-A2.
PD 03-OCT-2002.
PF 27-MAR-2002; 2002WO-IB002163.
XX
PR 27-MAR-2001; 2001GB-00007658.
XX
PA (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
PI Masignani V, Tettelin H, Fraser C;
XX
DR WPI; 2003-040579/03.
DR P-PSDB; ABU02123.
XX
PT New proteins and nucleic acid molecules from Streptococcus pneumoniae.
PT useful as medicaments for treating or preventing a disease or infection
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
XX ear infection.
PS Claim 6; SEQ ID NO 3399; 56pp; English.
XX
CC The invention relates to a protein comprising or having at least 50%
CC identity to any of the 2469 amino acid sequences, identified in the
CC specification (available on a computer readable format), or its fragment,
CC expressed from 2469 of 2489 identified DNA coding regions from the
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
CC AB556454. Also included are an antibody which binds one of the proteins,
CC treating a patient by administering the protein, DNA or antibody (in a
CC composition), a kit comprising first and second primers, which are the
CC nucleic acid cited above or fragments between nucleotides 8-100 of a

XX		The present invention relates to nucleic acid sequences (ABQ67188-
CC		ABQ71212) from Listeria sp. The sequences are useful as probes and
CC		primers for identification and/or detection of Listeria (e.g., as
CC		contaminants in foods, or mutational analysis) and for analysis of gene
CC		expression. Proteins encoded by the nucleic acid sequences can be used to
CC		screen for compounds that modulate gene expression, replication and
CC		pathogenicity of listeria (potential therapeutic agents), also for
CC		treating infections by Listeria, and are useful as immunogens in anti-
CC		Listeria vaccines. Note: The sequence data for this patent did not form
CC		part of the printed specification, but was obtained in electronic format
CC		directly from WIPO at ftp.wipo.int/pub/published_pct_sequences. (Updated
CC		on 29-AUG-2003 to standardise OS field)
XX		
SQ	Sequence 12278 BP; 4018 A; 2610 C; 1808 G; 3841 T; 0 U; 1 Other;	
	Query Match	12.3%; Score 75; DB 6; Length 12278;
	Best Local Similarity	46.5%; Pred.No. 2.3e-13;
	Matches 280; Conservative	0; Mismatches 316; Indels 6; Gaps 1;
QY	1	ATGATAGAAATCAATGACCTCAGAATACTTTTGGCGTTCGGATCTTTATGGCAAGTC TC 60
DB	9490	ATGATTGAATTAGTTAATGTGCATAAAAGATACAAGATAAATTGATTTTGAAAAANGTC 9431
QY	61	AGTCATTAAGTTTTTACCAGAAACAATGCACGACTGACCTGGAGCGTCCGGTTCAGGAAA 120
DB	9430	TCTCTTTTCTATTGGAGCTGGTGTAATTATTTCAGTCGTGGGGAGAGTGGTAGTGGGAAG 9371
QY	121	TCGACTTTTGCTCAACTCTCTTTGGCACACTTGACAAACCAAAGTTTCGGGACAGATCCCTTGTC 180
DB	9370	ACCAGCTGTAAATGTATTCGGACACCTAGATTCAAAGATAGTGGGAAGTATTATT 9311

[illegible]

RESULT 15

ACA39374
ID ACA39374 standard: DNA: 1401 bp.

ACA
ACA39374;

DT 19-JUN-2003 (first entry)

DE Prokaryotic essential gene #21031.

Antisense; ds; prokaryotic essential gene; cell proliferation; drug design; gene.

xx
OS Mycoplasma genitalium.

WO200277183-A2.

XX
PD
03-OCT-2002.

21-MAR-2002; 2002WO-US009107.

21-MAR-2001: 2001US-00815242

21-MAR-2001; 2001US-0081324Z;
06-SEP-2001; 2001US-00948993;
PR

PR 25-OCT-2001; 2001US-0342923P

PR 08-FEB-2002; 2002US-00072851..

PR 06-MAR-2002; 2002US-0362699P.

PA (ELIT-) ELITRA PHARM INC.

PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlson KL, Zyskind JW:

PI Wang D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH; Zyskind, Haseideck R, Olisen KL, Forsyth RA, Xu HH;

WPI: 2003-029926/02.

DR WEL; 2003-0299267
DR P-PSDB; ABU35504.

xx
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.

PS Claim 14; SEQ ID NO 27244; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published pct sequences

Sequence 1401 BP; 538 A; 201 C; 225 G; 437 T; 0 U; 0 Other;

Query Match 12.2%; Score 74.6; DB 7; Length 1401;

Best Local Similarity 50.1%; Pred. No. 1e-13;

Seq. Locat. Similarity 50.4%; Freq. NO. 18-15,
Matches 222; Conservative 0; Mismatches 209; Indels 12; Gaps 1;

QY 100 GGAGCGTCCGGTTCAGGAAATCGACTTTGCTCAACTGTCTTGGCACACTTGACAAACCA 159

Db 805 GGCCTTCAGGATCTGGTAAACCACATTACTAAACATTAATTCAGGGATGGATAGAGCT 864

OV 160 AGTTCCGGACAGATCCTTGTGTCGAGGATGTAGACCTTCTGAAACCTCTCTACCGCTAAGCAA 219

865 TCCTAGTGGTAGTGTTATTGTCAATGCTTATACATCATTCCTTAATCATTAACCTC
924

2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

[illegible]

0
10
20
30
40
50
60
70
80
90
100

DD T043 ATCAGGCTTGTATGTCACCTTTTAGAGAGCCGCTTGGGATGGATAGTTTGGCATTAGAGAGCTTCCCTTTTGT

QY 388 TTGAACCTCTGGTGGCGAACACAAAGAACTGCGTTGGCCGGGTACTGCTCAAAAAT 447

DD 1105 AAATGAAATGAGTGGTGGGCAACAGCAACGCTGTTCATATGCAAGAGCTTTTGTAAATAAC 1164

448 CCCGAAATAATCTGGCTGATGAACCAACCGGAGCTCTAGATTTAACAACAGTGAGCTA 507

Db 1165 CCCTTATTAAATTTTGGTGATGAACCTACTGGGGCACTTGATCTTGAGATGACCCAAATT 1224

Qy 508 GTCATAGAAGCATTTGAGAGCACT 530

Search completed: July 30, 2004, 12:27:05

us-09-868-338-7_copy_1117_1725.rng

Mon Aug 2 09:36:54 2004

Job time : 455 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 30, 2004, 10:28:19 ; Search time 99 Seconds
(without alignments)
3413.789 Million cell updates/sec

Title: US-09-868-338-7_COPY_1117_1725
Perfect score: 609
Sequence: 1 atgatagaataaatgaact.....cggataaccattatacaacta 609

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCFUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	75.4	12.4	21706	4	US-08-961-527-36
2	74.6	12.2	580073	4	US-08-545-528D-1
3	74.6	12.2	580073	4	US-08-545-528D-1
4	71.8	11.8	813	4	US-09-107-532A-3565
5	71.6	11.8	1080	4	US-09-543-681A-1259
6	71.2	11.7	771	4	US-09-134-001C-1763
7	69.6	11.4	681	4	US-09-134-000C-2474
8	68.4	11.2	11050	4	US-08-956-171E-96
9	66.6	10.9	1191	4	US-09-489-039A-3320
10	66	10.8	720	4	US-09-134-000C-2762
11	65	10.7	14872	4	US-08-961-527-72
12	64.8	10.6	711	4	US-09-134-001C-1758
13	64	10.5	1149	4	US-09-134-001C-602
14	63.8	10.5	4635	4	US-09-221-017B-187
15	62.6	10.3	4403765	3	US-09-103-840A-2
16	62.6	10.3	4411529	3	US-09-103-840A-1
17	61.6	10.1	765	4	US-09-134-001C-1603
18	61.6	10.1	813	4	US-09-134-001C-2724
19	61	10.0	6693	4	US-08-961-527-195
20	61	10.0	8114	4	US-09-453-702B-29
21	60.8	10.0	4659	2	US-08-583-276-18
22	60.8	10.0	20986	4	US-08-961-527-54
23	60.2	9.9	1830121	4	US-09-557-884-1
24	60.2	9.9	1830121	4	US-09-557-884-1
25	60	9.9	1221	4	US-09-643-990A-1
26	59.6	9.8	10690	4	US-08-961-527-93
27	59.2	9.7	2726	1	US-08-461-823-1

28	59.2	9.7	3988	4	US-09-762-195-1	Sequence 1, Appli
29	59.2	9.7	4284	2	US-08-784-649A-1	Sequence 1, Appli
30	59.2	9.7	4284	2	US-08-784-649A-5	Sequence 5, Appli
31	59.2	9.7	4646	1	US-08-181-471-2	Sequence 2, Appli
32	59.2	9.7	4646	4	US-09-023-655-1167	Sequence 1167, Ap
33	59.2	9.7	4669	6	5206352-3	Patent No. 5206352
34	59.2	9.7	6505	2	US-08-793-610-5	Sequence 5, Appli
35	59.2	9.7	8630	4	US-09-306-417-1	Sequence 1, Appli
36	59.2	9.7	8630	4	US-09-306-417-2	Sequence 2, Appli
37	59.2	9.7	9318	2	US-08-793-610-6	Sequence 6, Appli
38	58.8	9.7	666	4	US-09-107-532A-3013	Sequence 3013, Ap
39	58.8	9.7	666	4	US-09-134-000C-3112	Sequence 3112, Ap
40	58.8	9.7	3909	1	US-08-232-537-1	Sequence 1, Appli
41	58.6	9.6	771	4	US-09-107-532A-554	Sequence 554, App
42	58.6	9.6	8654	4	US-08-961-527-98	Sequence 98, Appl
43	58	9.5	774	4	US-09-134-000C-520	Sequence 520, App
44	57.6	9.5	3924	4	US-09-023-655-1168	Sequence 1168, Ap
45	57.6	9.5	3924	4	US-09-762-195-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-961-527-36
; Sequence 36, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21706 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-36

Query Match 12.4%; Score 75.4; DB 4; Length 21706;
Best Local Similarity 50.4%; Pred. No. 1.7e-15;
Matches 212; Conservative 0; Mismatches 206; Indels 3; Gaps 1;

OY 99 TGGAGCGTCCGGTTCAGGAAATCGATTGCTCACTGTTTCGCACACTTGACAAACC 158
DB 14961 TGGTCTTCAGTGCAGCAGTCAACAGTCTTAACTCTTGGGGGATGGATACAA 15020

159	QY	AA	GTTC	CGGACAGATCCTTGTGTCGAGGATGTAGACCTTCTGAAACTCTCTACGCGTAAGCA	218
15021	Db	TT	GAAGGGGAAATCTGGATTGATGGTGTAAATATTCCGGATTATAGTTCACACGACG	15080	
219	QY	AC	GGTTATACAGGAAATAACGGTGGGCTATTATTTTCAAGATTATGCCTTGATTCGGA	278	
15081	Db	CA	CAAAATACCGTAGAAATGATGTGGGTGTGTTTTTCAGTTTTATATACTAGTTTCTAA	15140	
279	QY	CAG	CACAGTTAAATTCACCTTTCAGCTTTCGGTGGAAAAACACAAATGGCGCTGAAATTC	338	
15141	Db	TTCT	GCACGCTAAGGAAATGTGGAATCTGGCTTCGAAATTTGTGCAGATGCTTGAATTC	15200	
339	QY	---	TTCAAGTACTTCATGCTGTGGTCTTGTAGTTCGTTTCGAGGAAAAAGCCAGTTTTTGAAC	395	
15201	Db	TT	CAGGCTTGACAGATGTAGTCTGGCTCATCGTCTCAATAACTTTCAGGCCAGCT	15260	
396	QY	CT	CGTGGCGGAAACACACAGNACTGCGTTGGCCGGGTACTGCTCATAAAATCCCCGAT	455	
15261	Db	TT	CTCGAGGGGAGCAACGCGAGTCTTCATTGCACGCGCGGTAGCCAAAAATCTCTAAAT	15320	
456	QY	AA	TCTCGCTGATGAACCAACCGGAGCTCTAGATTTAACAAACAGTAGTGCATAGA	515	
15321	Db	TTCT	CTTTGTGATGAACCGACTGGAGCCTTGGATTATCAGACGGGCAAGCAGGTTTGAA	15380	
516	QY	A	516		
15381	Db	A	15381		

RESULT 2

```

RECORD 1
US-08-545-528D-1
; Sequence 1, Application US/08545528D
; Patent No. 6537773
; GENERAL INFORMATION:
; APPLICANT: Fraser et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragment
; Patent No. 6537773
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PB193P1
; CURRENT APPLICATION NUMBER: US/08/545,528D
; CURRENT FILING DATE: 1995-10-19
; PRIOR APPLICATION NUMBER: US 08/488,018
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/473,545
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 580073
; TYPE: DNA
; ORGANISM: Mycoplasma genitalium
US-08-545-528D-1

```

328	CCTGAAATTCCTCAAGTACTCATGCTGTTGGTCTTGAGTCGTTTCGAGGAAAGCCAGTT	387
78730	ATCAGCACTGATGCACCTTTTAGAAGCGGTTGGGATGGATAGTTTGCAAAAGAGCTTCCT	78789
388	TTTGAACTCTCTGGTGGCGAACAACAGAACTCGGTTGGCCCGGTACTGCTCAAAAAT	447
78790	AATGAATTGAGTGGGCGCAACAGCAAGCTTTCCATTGCAAGAGCTTTTGCTTAAAC	78849
448	CCCCGAATAATTCGGCTGATGAACCAACCGAGCTCTAGATTTTAAACAAAGTGAAGCTA	507
78850	CCCTTATTAAATTTTGGTGATGAACCTACTGGGCACTTGATCTTGAGATGACCCAAAT	78909
508	GTCATAGAAGCATTCAGAGCACT	530
78910	GTTTAAACAGCTTTTAGCAAT	78932

QY
 Db
 QY
 Db
 QY
 Db
 QY
 Db

RESULT 3
 US-08-545-528D-1/c
 ; Sequence 1, Application US/08545528D
 ; Patent No.: 6537773
 ; GENERAL INFORMATION:
 ; APPLICANT: Fraser et al.
 ; TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragments
 ; Patent No. 6537773
 ; TITLE OF INVENTION: Thereof, and Uses Thereof
 ; FILE REFERENCE: PB193P1
 ; CURRENT APPLICATION NUMBER: US/08/545,528D
 ; CURRENT FILING DATE: 1995-10-19
 ; PRIOR APPLICATION NUMBER: US 08/488,018
 ; PRIOR FILING DATE: 1995-06-07
 ; PRIOR APPLICATION NUMBER: US 08/473,545
 ; PRIOR FILING DATE: 1995-06-07
 ; NUMBER OF SEQ ID NOS: 1
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 1
 ; LENGTH: 580073
 ; TYPE: DNA
 ; ORGANISM: Mycoplasma genitalium
 ; US-08-545-528D-1

QY 481 GCTCTAGATT 491
Db 576552 GCAGTAAATAT 576542

RESULT 4
US-09-107-532A-3565
; Sequence 3565, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Arinietello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3565:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 813 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (3) LOCATION 1...813
; SEQUENCE DESCRIPTION: SEQ ID NO: 3565:
US-09-107-532A-3565

Query Match 11.8%; Score 71.8; DB 4; Length 813;
Best Local Similarity 49.6%; Pred. No. 4.1e-15;
Matches 212; Conservative 0; Mismatches 212; Indels 3; Gaps 1;

QY 99 TGGAGCGTCGGTTTCAGGAAATCGACTTCTCAACTGCTTCTGAGGATGTAGACCTTCTGAAACTCTCTACGGGTAAAGCA 158
Db 228 TGGTCCAAAGTCGACGAGGAAGTCAACGATCTCTTAATTTTAGTGGGATGGATACACC 287

QY 159 AAGTTCGGGACAGATCCTTGTTCGAGGATGTAGACCTTCTGAAACTCTCTACGGGTAAAGCA 218
Db 288 TGATGAAGTCAGATCATTTATGTATGACAGATATTGCGCAATTTTCAGATTAACAGCT 347

QY 219 ACGGTTATACAGGAAATACGGTGGGCTATTTATTTCAAGATATGCTTGAATCCCGA 278
Db 348 TACTGCTTACCGACGACAGATGAGGGTTCGTTTCCCAATTTCTACAATCTCGTTCCTAA 407

QY 279 CAGGACAGTTAAATTCACCTTCAGCTTGGGTGGAAAAA---CACAAATGGCTGAAT 335
Db 408 TCTAACAGCAAAAGAAATGCGAATTCGCAACGGAAGTTTCTCTGATCGCTTGTATCC 467

QY 336 TCCTCAAGTACTTCATGCTGTTGCTTGTAGTGTTCGAGGAAAGCCAGTTTTCGAAT 395
Db 468 TGTAGAGGTATTGAGGCAAGTGGGCGCTAGCTCATCGATTGAATAACTTTCCGTCCTCAAT 527

QY 396 CTCTGGTGGGAAACAACAAGAACTCGTTGGCCGCGGTACTGCTCAAAAATCCCGAAT 455
Db 528 ATCAGAGGGGAACACAGAGAGGGTTTCAATTCGCTGCTAGCTAGCTAAATAATCCCAACT 587

QY 456 AATTCGCTGTATGAACCAACCGAGCTCTAGATTTTAAACAACAGTGTAGTCTATAGA 515
Db 588 TCTTTTATGTATGAACCGACAGCGCGCTGGATTTTGAACGGGCAACGAGTATTGAA 647

QY 516 AGCAATTG 522
Db 648 ATTATTG 654

RESULT 5
US-09-543-681A-1259
; Sequence 1259, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709,1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 1259
; LENGTH: 1080
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-1259

Query Match 11.8%; Score 71.6; DB 4; Length 1080;
Best Local Similarity 47.4%; Pred. No. 5.8e-15;
Matches 215; Conservative 0; Mismatches 239; Indels 0; Gaps 0;

QY 46 TTATGCAAGGTCTCAGTCATAAGTTTTTACAGGAACAATGACAGCTACTGCTGAGCG 105
Db 52 TTGTGCTCGGAAGTGAATACTCAATTACCAACAGAAAGTATCACGCGGTATTTGGCTTA 111

QY 106 TCGGTTTCAGGAAATTCGACTTTGCTCACTGCTTTGGCACACTTGACAAACCAAGTTCC 165
Db 112 TCTGGCGCAGGTAAACGTCGCTCATTAATGTGATTTGGTTTAAACCAACCTGATTA 171

QY 166 GGACAGATCCTTTGTCGAGATGTAGACCTTCTGAAACTCTCTACCGTAAGCAACGGTTA 225
Db 172 GGACATATTGTGTTAAATAATCACACATTTGGTGGATATTGAGAAGAAAAATTTATCTGCCA 231

QY 226 TACAGGAAAAATACGCTGGCTTATTTTTCAGATTTATGCTTGAATTCCTGATCCGACAGCA 285
Db 232 CCGGAAAGCGCAAGTAGGCTATGCTATTTCAAGATGCACACTATTTTCTCTATTATACC 291

QY 286 GTTAAATTCACCTTCAGCTTCGGTGGGATGAAAAACAACAATGGCTGAAATTCCTCAAGTA 345
Db 292 GTAAAGGAAATTTACTTTAGCTATGTCCACCAATGAAAGTACAGTTTGTATGAAT 351

QY 346 CTTTCATGCTGTGGTCTTTCAGTTCGTCAGGAAAGCCAGTTTTCAGTCTCTGCTGCGC 405
Db 352 ATTTATCTATTAGGTATGCAACATCTACTTTTCGGGTTTTCCTATTACGCTATCTGCTGCT 411

QY 406 GAACAACAACGAACTGCTGCTGGCCGGGTACTGCTCAAAAATCCCGCAATAATTCGCTGCT 465
Db 412 GAACAACAACGCTGCTGCTTGGAGTCAATTAATTAACGACCCGATATTATGCTCAATG 471

```
QY 466 GATGAACCAACCGGAGCTCTAGATTAAACAACA 499
Db 472 GATGAACCAATGGCGCTCTCTGATTACCCGCA 505

RESULT 6
US-09-134-001C-1763
; Sequence 1763, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1763
; LENGTH: 771
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1763

Query Match 11.7%; Score 71.2; DB 4; Length 771;
Best Local Similarity 49.6%; Pred. No. 6.5e-15;
Matches 227; Conservative 0; Mismatches 213; Indels 18; Gaps 1;

QY 57 TCTCAGTCATAGTTTATACAGGAACATGACACACTGAGCGTCCGGTTCAGG 116
Db 84 TATAAGTTTTCAGTTGAAAGGGGAATTTGAGCTGTCTGATGGTCTCTGGATCAGG 143

QY 117 AAAATCGACTTTGCTCAACTGCTTGGCACACTTCAACAAACCAAGTTCCGGACAGATCCT 176
Db 144 TAAACGAGCTTTTACTCAATGTGTAAAGTTCAATAGATATCTATTTCCAGGAGGAAGTGTAGA 203

QY 177 TCTCGAGGATGAGACCTTCTGAACTCTACGGTGAAGCAACGGTTATACAGAAAAA 236
Db 204 AGTTGAAGCAAGAAATTAATAAAGTGAAGCAAGAGTGGCAATTTTCGAAACA 263

QY 237 TACGGTGGGCTATTATTTCAAGATTATGCTTGAATCCCGACAGGACAGTTAAATTCOA 296
Db 264 ACATCTCGGTTTATTTTCAAGATTATAGCGTTTACCCACATTAAACAGTAAAGAAAA 323

QY 297 CCTTCAGCTTGGGTGGAAACACAAATGGGCTGAATTCCTC----- 340
Db 324 TATTATGCTACCACTCTAGTACAAAAATTCATAAATATGAATGGAACAAATTAATA 383

QY 341 --AAGTACTTTCATGCTTGGTCTTGGTCTGAGTCGTTCCAGGAAAGCCAGTTTGTGAACCTC 398
Db 384 AGAAGTGGCTGAGCAATTAGTATTTATAAAGTGGAAATTAATATCCAAAGTGAATTTTC 443

QY 399 TGGTGGCAACAAACAGAACTGCTTGGCGGGTACTGCTCAAAATCCCGCAATAT 458
Db 444 TGGCGGTGACCAACAGTACGCGGCGCGCGGCAATTCGTCATAAACCAACAGGATTAT 503

QY 459 TCTGGCTGATGAACCAACCGGAGCTCTAGATTAAACA 496
Db 504 TTTCGAGATGAACCTACTTGGCGCATTAATTTCTAAA 541

RESULT 7
US-09-134-000C-2474
; Sequence 2474, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: GTC-007
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1763
; LENGTH: 771
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1763
```

```
FILE REFERENCE: 032796-032
CURRENT APPLICATION NUMBER: US/09/134,000C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/055,778
PRIOR FILING DATE: 1997-08-15
NUMBER OF SEQ ID NOS: 6812
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2474
LENGTH: 681
TYPE: DNA
ORGANISM: Enterococcus faecalis
US-09-134-000C-2474

Query Match 11.4%; Score 69.6; DB 4; Length 681;
Best Local Similarity 49.4%; Pred. No. 2.2e-14;
Matches 251; Conservative 0; Mismatches 239; Indels 18; Gaps 2;

QY 77 CAGGAACAATGACAGCACTGACTGGAGCTCGGGTTCAGGAAATCGACTTTGCTCAACT 136
Db 95 CTGGAGAATTTTTCAGGATTTATGGGTCTCTCTGGTTCGGGAAATCGACTTTGATTAAT 154

QY 137 GTCCTGGCACACTTGCACAAACCAAGTTCGGGACAGATCCTTGTCCAGGATGAGACCTTC 196
Db 155 TGCTAGGCTTTATTCATAAAGTTTGAAGGAACCTATTATTATTGAAGACCGTGAATTTG 214

QY 197 TGAACCTCTACGGGTAAAGCAACGGTTATACAGGAAATACGGTGGGCTATTATTATTC 256
Db 215 GCGACTTTTCTGATAAGGAACCTATCCCGAATTTAGAAATGAGCAGTCGGCTTTGCTTTTC 274

QY 257 AAGATTATCGCTTGTATCCGACAGGACAGTAAATTTCAACCTTCAGCTTTCGGGT----- 311
Db 275 AGAATTTTGTATTTTAAATGAAACACTAACAGTTGAGAAATACATTGAATTTACCTCTTTAT 334

QY 312 -----GGAAAAACAATGCGCTGAAA-----TTCCTCAAGTACTTCAATGCTGTG 358
Db 335 ATAGTGGGTTAAACCCCAAGGAAGCAAGATCGTGTCCACGAAGTTCTGACAAAAGTCG 394

QY 359 GCTCTGAGTCGTTCCGAGGAAAGCCAGTTTGTGAACCTCTCTGTCGCGAACAACAACGAA 418
Db 395 GCTTGCAGATAAAGCAAGAGATCCAAACACATTTTCGGCGGACACAAACACGTCG 454

QY 419 CTGCGTTGGCCCGGTPACTGCTCAAAATATCCCGAATTAATCTGGCTGATGAAACACCG 478
Db 455 TTCCCATTCGGGTGGGATTTGTCATTCGCCCAAGTTTATTATTGACAGATGAACCACTG 514

QY 479 GAGCTTAGATTAAACAAACAGTGAAGTACTAGTATGATGAAGCATTGAGAGCACTCGCGACA 538
Db 515 GGCCTTTAGATAGTAAACATCGGAAGAAATTTTAAGCTCTTTTCAGCAATTAACAATG 574

QY 539 AAGCGCCACCGTTGTTGTTGCTACGCA 566
Db 575 AAGGTGTCACCATTTATTTAGTACACA 602

RESULT 8
US-08-956-171E-96
; Sequence 96, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
```



```

RESULT 10
US-09-134-000C-2762
; Sequence 2762, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2762
; LENGTH: 720

```

; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-134-000C-2762

Query Match 10.8%; Score 66; DB 4; Length 720;
Best Local Similarity 55.8%; Pred. No. 4.4e-13; Indels 0; Gaps 0;
Matches 126; Conservative 0; Mismatches 100;

QY 341 AAGTACTTTCATGCTGCTTGGTCTCTGAGTCGTTTCGAGGAAAGCCAGTTTTCGAACTCTCTG 400
DB 413 AATACCTAAATAATGTCGGTTTGGAAACAAAGCCAAATCAAAAGTTACTGAACTTCTG 472

QY 401 GTGGCGAACACAGCACTGCTGTTGGCCGGTACTGCTCAAAATCCCGAATAATTC 460
DB 473 GAGGACACCAACAAAGAGTAGCCATTGCTGCTTCTTGGTGAATGATCCATTATGATTA 532

QY 461 TGGCTGATGAACCAACCGGAGCTCTAGATTTAAACAACAGTGAAGTCTAGTCATAGAAGCAT 520
DB 533 TGGCTGATGAACCGACTGGGCAATGGATAGTGAACGGGTACAGAATATATGAATAT 592

QY 521 TGAGAGCACTCGCCACAAAGGCGCCACCGTTGTTGTTGCTACGCA 566
DB 593 TGAAGAATTGAATGAACAGGGCAAAACAATGTGATGTCATGTCACACA 638

RESULT 11
US-08-961-527-72
; Sequence 72, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 72:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14872 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-72

Query Match 10.7%; Score 65; DB 4; Length 14872;
Best Local Similarity 49.1%; Pred. No. 6.7e-12; Indels 18; Gaps 2;
Matches 244; Conservative 0; Mismatches 235;

QY 91 GCAGTACTGGAGCGTCGGTTTCAGGAAATTCGACTTTGCTTCAACTGCTTGGCACACTT 150

DB 207 GCCATCATGGGTGAGTCTGGTTCTTGGTAAATCAACTCTTCTCAATATTTCTAGCTAATGTTG 266
QY 151 GACAAACCAAGTTCGGGACAGATCTTGTGCGAGGATGTAGACCTTCTGAAAACCTCTCTACG 210
DB 267 GATAAACCAAGTCGGTGGTCAGGTTTACITTGATGAACTGACACCGCAACTATTAAAT 326
QY 211 CGTAAGCAACGGTTATACAGGAAATACGGTGGGCTATTATTATTCAGATATTCCTTGG 270
DB 327 TCACAGGCTTCTAGTTTTCGGGCGTGAAAAGCTAGGATTTGTCTTCCAAGACTTTAACTTG 386
QY 271 ATTCCGACAGGACAGTGTAAATTTCAACCTTCAGCTTGGGTGGTGGAAAAACACAAATGGCCT 330
DB 387 CTAGATACTCTGCTGTGTGTAAGGACAAATATCTTGGCTTCCGCTTGTCTTGCAGAGACT 446
QY 331 GAAAT-----TCTCAAGTACTTTCATGCTG-----TTGGTCTTGAAGTCGTTTC 372
DB 447 ATAACGGAGATGATGAAGAAATTTGGTGTGACAGCTGAGAACTCTGGGTATTAAACCAATTG 506
QY 373 GAGGAAACCCAGTTTGTGAACCTCTCTGTGGCGAACACACAGAACTGCGTTGGCCCGG 432
DB 507 CAAGAGAGTACCCCTTACGAGATTTCTGTGTGTGAGAAACCGGTGTAGCAGTAGCCCGC 566
QY 433 GTACTGCTCAAAATCCCGGAATAATTCTGGCTGATGAACCAACCGAGCTCTAGATTTA 492
DB 567 GCCATCATCAGAAACCTGAAATTTCTCTTGGGACGAGCCAAACAGGAGCCCTTGATTC 626
QY 493 ACAACAGTGAAGTCTAGTCATAGAACATTCAGCACTGCGCGACAAAGGCGCCACCGTT 552
DB 627 AAGTCATCTGCAGCCTTACTTGATGCTTTAATGAATCAATGAGCGTGGGCAACCAATC 686
QY 553 GTTGTGCTACGCACTC 569
DB 687 CTCATGTAACCACTC 703

RESULT 12
US-09-134-001C-1758
; Sequence 1758, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1758
; LENGTH: 711
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1758

Query Match 10.6%; Score 64.8; DB 4; Length 711;
Best Local Similarity 47.2%; Pred. No. 1.2e-12;
Matches 250; Conservative 0; Mismatches 262; Indels 18; Gaps 1;

QY 47 TATGCGAAGTCTCAGTCAATAGTTTTCACGAGAACATGACAGCACTGACTGAGCGT 106
DB 83 TTTTGAAGACATAGATATTCATATGATGAGGTGAATTTATTGCTATCATGTCGTCG 142
QY 107 CCGGTTTCAGGAAATCGACTTTGCTTCAACTGTCTTGGCACACTTTCACAAACCAAGTTCCG 166
DB 143 CTGGTTCAGGTAAGTAAGTACGTTAATTAATCTTGGGATTTATTGATCGTGGATATGA 202
QY 167 GACAGATCCTTGTGCGAGATGTAGACCTTCTGAACTCTTACCGGTAAAGCAACGGTTAT 226
DB 203 GAGAATACTTTTTTAAACAATGAGATTTATCAAAAAGCTCAGATAATTAAGCTCCGAGAAA 262

QY 227 ACAGGAAATACGGTGGCTATTATTTCAAGATTATCCCTTGATTCCTCCGACAGACAG 286
Db 263 TTCCGAATCATACTGTAGTGTTCGTCTTTCCAGAAATTTTAAATTTGATTCAAAATAACACTA 322
QY 287 TTAATTTCAACTTTCAGCTTGGGTGG-----AAAAACAACAATGCG 328
Db 323 TTTTGAATAATGTAAGTATTCGCTAAATTTAATGTTTGTAGTAAATAAGCGAGAAA 382
QY 329 CTGAAATTCCTCAAGTACTTATGCTGTGTGGTCTTGAGTCTGTCGAGGAAAGCGAGTTT 388
Db 383 GCAAGTCTTTAGACGACTTCATGCTCGGTCTAAAGGTTAAAGAGAAATCTCTTACCAA 442
QY 389 TTCAACTCTCTGGTGGGACACACACAGTGGCTGGTGGCGGCTGCTCAAAATC 448
Db 443 ATAAATTTGTCAGAGGGGACACACAGTGTAGCTATTGCCCGGGCAATTAATATGACC 502
QY 449 CCCGAATAATTCCTGGTGTATGAACCAACCGGAGCTCTAGATTTAAACACAGTGTAG 508
Db 503 CTAAGTTTATTTGCGGTGACCTTCTGAGTCTGCGCTCTTGAATCGAAGACATCTCAAGATA 562
QY 509 TCATAGAGCTTGTAGAGACTCGCGCAACAGGCGCCACCGTGTGTGTT 558
Db 563 TTATGAGCTTTTCGTAAACTGAACAAAGACAGAAATCAACTATATT 612

RESULT 13
US-09-134-001C-602
; Sequence 602, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 602
; LENGTH: 1149
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-602

Query Match 10.5%; Score 64; DB 4; Length 1149;
Best Local Similarity 45.2%; Pred. No. 3e-12;
Matches 235; Conservative 0; Mismatches 285; Indels 0; Gaps 0;

QY 11 TCATGACCTCAAGAAATCTTTGGCGTTCGGATCTTATGGCAAGTCTCAGTCATAAGT 70
Db 74 TTAAGATGTCAGTAGGGCTTTGAAGATGTACAATACTAAATGAAATTAATATTGATA 133
QY 71 TTTTACAGAAACAATGACAGCTACTGAGCGTTCGGCTTCCAGTAAATCGACTTTGC 130
Db 134 TTGAACCCAGGCTATTTTATACACTATTAGTCCCTCAGGTTGTGGAACCAACAATTT 193
QY 131 TCAACTGCTGGCACACTTGACAAACCAAGTTCCGACAGATCTTGTGAGGATGTAG 190
Db 194 TAAACTCATAGCAGGATTTGAATATCCGATAGTGAGATATATATAAGATAAAC 253
QY 191 ACCTTCTGAACTCTCTACGCGTAAAGCAACGGTTATACAGGAAATAACGGTGGGTATT 250
Db 254 CTAATGTTAAATGCCACCAGTAAGGTAAGGTAAGTAAATCTGATTCCAGACTATGCAT 313
QY 251 TATTTCAAGATTATGCTTCCGACAGGACAGTAAATTCACCTTCAGCTTCCGG 310
Db 314 TGTTTCCACATTTAAATGTATTCGACAATATGCTATGTTTAAATTAATAAATTA 373
QY 311 TGAACAAACAAATGGCCTGAATTCCTCAAGTACTTCAATGCTGTGTTGGTCTTGAAGTCGT 370

Db 374 GTAAGTCAGAAATTAACGGTAAGGTTACTGAAGCACTTCAGTTGGTGAATTAAGTGGTT 433
QY 371 TCGAGGAAAAAGCCAGTTTTTTGAACTCTCTGTCGCGGAAACAAACAGAACTGCGTTGCCCC 430
Db 434 ATGAACATAGGCAATACAAGGTATGAGTGTGGACAAAAACACACGTTGTAGCCATAGCAC 493
QY 431 GGGTACTGCTCAAAATCCCGAATATTTCTGGCTGATGAACCAACCGGAGCTCTAGATT 490
Db 494 GGGCAATGTTATGAGCCGAAATATTTATTAGATGAGTCTTTATCGCATTAGATT 553
QY 491 TAAACAAACAGTGTAGCTAGTCTATGAGCAATTCATGAGCACTTGTAGAGCACT 530
Db 554 TAAATTAACGAACCTGAAATGCAATATTTATTGAGAGAACT 593

RESULT 14
US-09-221-017B-187/c
; Sequence 187, Application US/09221017B
; Patent No. 644799
; GENERAL INFORMATION:
; APPLICANT: Ross, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA: US/09/221,017B
; FILING DATE: 23-DEC-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1182
; FILING DATE: 31-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1546
; FILING DATE: 30-JAN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP2911
; FILING DATE: 09-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AUS98/01023
; FILING DATE: 10-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Morley, Gladys H
; REGISTRATION NUMBER: 32,430
; REFERENCE/DOCKET NUMBER: 27340-20021.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-813-5600
; TELEFAX: 650-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 187:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4635 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: UNKNOWN
; ORIGINAL SOURCE:
; ORGANISM: PORYPHYROMONAS GINGIVALIS
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1...4635

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 30, 2004, 13:27:50 ; Search time 506 Seconds
(without alignments)
5901.210 Million cell updates/sec

Title: US-09-868-338-7_COPY_1117_1725

Perfect score: 609
Sequence: 1 atgatagaatcaatgacct.....cggataccattatcaacta 609

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3222919 seqs, 2451570024 residues

Total number of hits satisfying chosen parameters: 6445838

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA.*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
- 17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 18: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
- 19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	80.8	13.3	684707	16 US-10-398-221-9	Sequence 9, Appli
c	80.8	13.3	3011208	16 US-10-398-221-2058	Sequence 2058, Ap
2	77.4	12.7	711	9 US-09-815-242-9384	Sequence 9384, Ap
3	75.4	12.4	21706	13 US-10-158-844-36	Sequence 36, Appl
c	75	12.3	12278	16 US-10-398-221-3886	Sequence 3886, Ap
4	74.6	12.2	1401	13 US-10-282-122A-27244	Sequence 27244, A
5	74.6	12.2	31241	10 US-09-960-870-13	Sequence 13, Appl
6	74.6	12.2	31241	10 US-09-960-858-13	Sequence 13, Appl
7	74.6	12.2	31241	15 US-10-251-668-13	Sequence 13, Appl
8	74.6	12.2	580073	15 US-10-205-220-1	Sequence 1, Appli
c	74.6	12.2	580073	15 US-10-205-220-1	Sequence 1, Appli
9	74.6	12.2	756	13 US-10-282-122A-34668	Sequence 34668, A
10	73.6	12.1	5312	16 US-10-398-221-3820	Sequence 3820, Ap
11	73.2	12.0	684	13 US-10-282-122A-30817	Sequence 30817, A

15	72.4	11.9	666	9 US-09-738-626-956	Sequence 956, App
16	72.4	11.9	789	13 US-10-627-476-441	Sequence 441, App
c	72.4	11.9	3309400	9 US-09-738-626-1	Sequence 1, Appli
17	71.6	11.8	1062	13 US-10-282-122A-32942	Sequence 32942, A
18	71.2	11.7	319630	16 US-10-398-221-7	Sequence 7, Appli
19	69.6	11.4	32768	9 US-09-070-927A-123	Sequence 123, App
20	68.6	11.3	630	15 US-10-307-234-12	Sequence 12, Appl
21	68.6	11.3	7000	15 US-10-307-234-3	Sequence 3, Appli
22	68.6	11.3	11050	8 US-08-781-986A-96	Sequence 96, Appl
23	68.4	11.2	11050	13 US-10-329-624-96	Sequence 6038, Ap
24	68.4	11.2	687	9 US-09-815-242-6038	Sequence 238, App
25	68.2	11.2	29729	9 US-09-070-927A-238	Sequence 10135, A
26	65.6	10.8	1149	13 US-10-282-122A-10135	Sequence 27896, A
27	65.2	10.7	1398	13 US-10-282-122A-27896	Sequence 24170, A
28	65	10.7	726	13 US-10-282-122A-24170	Sequence 3721, Ap
29	65	10.7	2293	16 US-10-398-221-3721	Sequence 72, Appl
30	65	10.7	14872	13 US-10-158-844-72	Sequence 68, Appl
31	64.6	10.6	8921	9 US-09-070-927A-68	Sequence 187, App
c	63.8	10.5	4635	13 US-10-194-163-187	Sequence 1791, Ap
32	63.6	10.4	732	16 US-10-398-221-1791	Sequence 6, Appli
c	63.6	10.4	213251	16 US-10-398-221-6	Sequence 2058, Ap
33	63	10.3	3011208	16 US-10-398-221-2058	Sequence 9252, Ap
34	62.6	10.3	1163020	16 US-10-398-221-10	Sequence 3588, Ap
35	62.2	10.2	816	9 US-09-815-242-9252	Sequence 2421, Ap
36	62.2	10.2	816	9 US-09-815-242-9592	Sequence 195, App
c	62	10.2	3438	16 US-10-398-221-3588	Sequence 29, Appl
37	61.6	10.1	1059	13 US-10-282-122A-23957	Sequence 342, App
38	61.4	10.1	577	9 US-09-974-300-2421	
39	61	10.0	6693	13 US-10-158-844-195	
c	61	10.0	8114	15 US-10-114-170-29	
c	60.8	10.0	741	10 US-09-769-787-342	

ALIGNMENTS

RESULT 1

US-10-398-221-9
; Sequence 9, Application US/10398221
; Publication No. US2004001851441
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; PRIOR FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 684707
; TYPE: DNA
; ORGANISM: Listeria innocua
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(end)
; OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u
US-10-398-221-9

Query Match 13.3%; Score 80.8; DB 16; Length 684707;
Best Local Similarity 47.2%; Pred. No. 5.3e-14;
Matches 284; Conservative 0; Mismatches 312; Indels 6; Gaps 1;

QY	1	ATGATCAAAATCAATGACCTCAAGAAATCTTTGGGCTTCGGATCTTATCGCAGGTCTC	60
Db	346005	ATGATCGAATAGTAAATGTCAGTAAAGATCAAGATAAATTGATTTGGAAAAGTC	346064
QY	61	AGTCATAAGTTTATACAGGAACAATGACAGCACTGACTCGCGCTTCAGGAAA	120

Db 346065 TCCTCTTCTATTGGAGCTGGCGAATTTATTGACGTCTGCGGAGAGTGGTAGTGGGAAG 346124
QY 121 TCGACTTTCTCACTGCTTGGCACACTTGGCACAAACCAAGTTCGGACAGATCCTTGTG 180
Db 346125 ACAACGCTGCTAAATGTTATCGGACACTTGAATTCAGAAATAGTGGCAAGTTATTT 2934630
QY 181 GAGGATGTAGACCTTCTGAACTCTCTACGCGTAAAGCAACGGTTATACAGGAAATAACG 240
Db 346185 AACGAGATGGAATATCAGA-----CGAAAAAGAGGTTATGACTCTAATAAAGAGGTG 2934576
QY 241 GTGGGCTATTATTCAAGATTATGCTTGCCTTGCCTCCGACAGGACAGTTAAATTCACCTT 300
Db 346239 TTAGGTTTATATTCCAAAATTAICTATTGATGAAATGAAACAGTGTGAAACTTA 2934516
QY 301 CAGCTTGGCGTGGAAAAACACAAATGGCCTGAAATTCCTCAAGTACTTTCATGCTGTTGGT 360
Db 346299 TCATTACAGGTGGGGAATCGCAAGCTGATGATAGAGCATTTGGAAGAGTAGGAATG 2934456
QY 361 CTGAGTCTCTCGAGAAAGCCAGTTTTCAGCTCTCTGTCGCGAAACCAACCACT 420
Db 346359 GATGAGACTATTAGCAAAAAAGATATACCAATTAAGTGGTGAGAAAAACAACGGATT 2934396
QY 421 GCGTTGGCGCGGTACTGCTCAAAATCCCGAATTAATTCGGTGTGATGAACCAACCGGA 480
Db 346419 GCTATTGCGCATTTTACTCAACCATTTCACTTTTCTGCGGACCAACTGGC 2934336
QY 481 GCTCTAGATTAAACACAGTGGAGTCTAGTATGAGCACTTCGCGACAAA 540
Db 346479 AATTAGATGATAAAACAAACAAATCATTTGAATTAATTTCTAGCTTGAAGAACAA 2934276
QY 541 GGCGCCACCGTGTGTTGCTACGCACTGGCCCTCTTCCGAGAAATCAGCGGATACCATT 600
Db 346539 GGTAAGACTACTGTTGTGTGTCAGCATGACCGGAAATATCAGGAAAGCAGATCGGGTC 2934216
QY 601 AT 602
Db 346599 AT 346600

RESULT 2

US-10-398-221-2058/c
; Sequence 2058, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; PRIOR FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2058
; LENGTH: 3011208
; TYPE: DNA
; ORGANISM: Listeria innocua
US-10-398-221-2058

Query Match 13.3%; Score 80.8; DB 16; Length 3011208;
Best Local Similarity 47.2%; Pred. No. 1.2e-13;
Matches 284; Conservative 0; Mismatches 312; Indels 6; Gaps 1;
QY 1 ATGATAGAAATCAATGACTCAAGAAATCTTTGGCGTTCGGATCTTATGGCAAGTCTC 60
Db 2934809 ATGATTGAATAGTAAATGTGATGATAAAGATACAGATAAATGATTTTGGAAAGATC 2934750
QY 61 AGTCATAAGTTTTTACCAGGAACATGACAGCACTGACTGAGCGTCCGGTTCAGGAAA 120
Db 2934749 TCTCTTCTATTGGAGCTGGCGAATTTATTTCAGTCTGCGGAGAGTGGTAGTGGGAAG 2934690

RESULT 3

US-09-815-242-9384
; Sequence 9384, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9384

QY 121 TCGACTTTCTCACTGCTTGGCACACTTGGCACAAACCAAGTTCGGACAGATCCTTGTG 180
Db 2934689 ACAACGCTGCTAAATGTTATCGGACACTTGAATTCAGAAATAGTGGCAAGTTATTT 2934630
QY 181 GAGGATGTAGACCTTCTGAACTCTCTACGCGTAAAGCAACGGTTATACAGGAAATAACG 240
Db 2934629 AACGAGATGGAATATCAGA-----CGAAAAAGAGGTTATGACTCTAATAAAGAGGTG 2934576
QY 241 GTGGGCTATTATTCAAGATTATGCTTGCCTTGCCTCCGACAGGACAGTTAAATTCACCTT 300
Db 2934575 TTAGGTTTATATTCCAAAATTAICTATTGATGAAATGAAACAGTGTGAAACTTA 2934516
QY 301 CAGCTTGGCGTGGAAAAACACAAATGGCCTGAAATTCCTCAAGTACTTTCATGCTGTTGGT 360
Db 2934515 TCATTACAGGTGGGGAATTCGCAAGCTGATGATAGAGCATTTGGAAGAGTAGGAATG 2934456
QY 361 CTGAGTCTCTCGAGAAAGCCAGTTTTCAGCTCTCTGTCGCGAAACCAACCACT 420
Db 2934455 GATGAGACTATTAGCAAAAAAGATATACCAATTAAGTGGTGAGAAAAACAACGGATT 2934396
QY 421 GCGTTGGCGCGGTACTGCTCAAAATCCCGAATTAATTCGGTGTGATGAACCAACCGGA 480
Db 2934395 GCTATTGCGCATTTTACTCAACCATTTCACTTTTCTGCGGACCAACTGGC 2934336
QY 481 GCTCTAGATTAAACACAGTGGAGTCTAGTATGAGCACTTCGCGACAAA 540
Db 2934335 AATTAGATGATAAAACAAACAAATCATTTGAATTAATTTCTAGCTTGAAGAACAA 2934276
QY 541 GGCGCCACCGTGTGTTGCTACGCACTGGCCCTCTTCCGAGAAATCAGCGGATACCATT 600
Db 2934275 GGTAAGACTACTGTTGTGTGTCAGCATGACCGGAAATATCAGGAAAGCAGATCGGGTC 2934216
QY 601 AT 602
Db 2934215 AT 2934214

```

; LENGTH: 711
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(711)
US-09-815-242-9384

Query Match      12.7%; Score 77.4; DB 9; Length 711;
Best Local Similarity 50.8%; Pred. No. 1.4e-14;
Matches 236; Conservative 0; Mismatches 226; Indels 3; Gaps 2;

Qy 55 GGTCTCAGTCATAGTTTACACAGGAACAATGACAGCACTGACTGGAGCGTCGGTTCA 114
Db 82 GATCTGAATTTGAGATTGAAGGGGAACTGGTTATTATCTTGGTCTTCAGGTGCA 141
Qy 115 GGAATAATCGACTTTGCTCAACTGCTTTGGCACACTTGACAAACCAAGTTCCGACAGATC 174
Db 142 GCGAAGTCAACAGTTCTTAACCTTCTTGGGGGAATGATACCAATGATGAAGGGGAATC 201
Qy 175 CTTCGCGAGATGAGACCTTCTGAACTCTACGCGTAAGCAACGGTTTATACAGGAA 234
Db 202 TGGATTGATGGTGTATATATTGCGGAATATAGTTCCACGCGCAACCAATTACCGTAGA 261
Qy 235 AATACGGTGGGCTATTTATTTCAGATTATGCGCTTGATCCCGACAGACAGTTAAATTC 294
Db 262 AATGATGGGGTTGTTTTTTCAGTTTATAATCTAGTTTCTAATCTGACAGCTAAGGAA 321
Qy 295 AACCTTCAGCTTGGGTGGGAAA--AACACAATGGGCTGAAAT--TCCTCAAGTACTTCAT 351
Db 322 AATGTGAACCTGGCTTCTGAAATTTGTACAGATGCGCTTGAATTTCTGATCAGGCTTTGACA 381
Qy 352 GCTTTGGCTTGTAGTCTGTCGAGGAAAGCCAGTTTGTGAATCTCTGCTGGCGGAACAA 411
Db 382 GATGTAGTCTGGCTCATCTGCTAAATACTTTCCAGCCAGCTTCTTGAGGGAGCAA 441
Qy 412 CAACGAATCGTTGGCGGGTACTGCTCAAAATCCCGAATAAATCTGGCTGTATGAA 471
Db 442 CAGCGAGTCTCATTGTCACGCGGTAGCCAAAATCTTAAATTTCTCTTTGTGTGAA 501
Qy 472 CCACCGGAGCTCTAGATTTTAAACAACAGTGAAGTCTAGTATAGAA 516
Db 502 CCGACTGGAGCTTGGATTATACACGCGGCAAGCAGGTTTGGAA 546

RESULT 4
US-10-158-844-36
; Sequence 36, Application US/10158844
; Publication No. US20040029118A1
; GENERAL INFORMATION:
; APPLICANT: Kunsch et al.
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-R
; COMPUTER: Dell Latitude Pentium 3
; OPERATING SYSTEM: Windows 98
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/158,844
; FILING DATE: 03-Jun-2002
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/961,527
; FILING DATE: 1997-10-30
; APPLICATION NUMBER: US 60/029,960

```

```

; FILING DATE: 1996-10-31
; ATTORNEY/AGENT INFORMATION:
; NAME: Hyman, Mark J.
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB340P1D1
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21706 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 36:
US-10-158-844-36

Query Match      12.4%; Score 75.4; DB 13; Length 21706;
Best Local Similarity 50.4%; Pred. No. 4.5e-13;
Matches 212; Conservative 0; Mismatches 206; Indels 3; Gaps 1;

Qy 99 TGGAGCGTCCGGTTTCAGGAAAATCCACTTTGCTCAACTGTCTTGGCACACTTGACAAACC 158
Db 14961 TGGTGCTTCAGGTGCGAGCAAGTCAACAGTTCTTTAACTTCTTGGGGAATGGATACCAA 15020
Qy 159 AAGTTCGGGACAGATCTTGTGCGAGGATGAGACCTTCTGAACTCTCTACGCGTAAGCA 218
Db 15021 TGATGAAGGGGAAATCTGGAATGATGGTGTAAATATTCGGGATTTAGTTCCACCGCG 15080
Qy 219 ACGGTTATACAGAAAATACGGTGGGCTATTATTTCAAGATTATGCTTGAATTCCTCGGA 278
Db 15081 CACCAATTACCGTAGAATGATGCGGGTGTGTTTTCAGTTTATAATCTAGTTTCTAA 15140
Qy 279 CAGGACAGTTAAATTCACCTTCAGCTTGGGTGGGTGGAACACAAATGGCTGAAATTC 338
Db 15141 TCTGACAGTAAAGGAAAATGTGAACTGGCTTCTGAAATTTGTGACAGATGCTTGAATCC 15200
Qy 339 ---TCAAGTACTTTCATGCTGTTGTTGCTTTCAGTTCGTCGAGGAAAAGCCAGTTTGAAC 395
Db 15201 TGATCAGGCTTTCAGATGATGAGTCTGCTCATCGTCTCAATACTTTCAGCCCGAGT 15260
Qy 396 CTCTGGTGGCGAACAAACAGAACTGCGTTGGCCGGGTACTGCTCAAAAATTCCTCGGAA 455
Db 15261 TTCTGGAGGGAGCAACAGCGAGTCTCCATTGCAAGCGGTAGCCAAAATTCCTAAAT 15320
Qy 456 AATTCTGGCTGATGAACCAACCGGAGCTCTAGATTTTACAAACAGCTGAGTATAGTA 515
Db 15321 TCTCCTTTGTGATGAACCGACTGGAGCTTGGATTATCAGACGGGCAAGCAGGTTTGA 15380
Qy 516 A 516
Db 15381 A 15381

RESULT 5
US-10-398-221-3886/c
; Sequence 3886, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3886
; LENGTH: 12278
; TYPE: DNA
; ORGANISM: Listeria monocytogenes 4b
; FEATURE:

```

```
; NAME/KEY: misc_feature
; LOCATION: (1)..(end)
; OTHER INFORMATION: n can be any nucleotide: a,g,c or t/u
US-10-398-221-3886

Query Match      12.3%; Score 75; DB 16; Length 12278;
Best Local Similarity 46.5%; Pred. No. 4.4e-13;
Matches 280; Conservative 0; Mismatches 316; Indels 6; Gaps 1;

QY   1 ATGATAGAAATCAATGACCTCAAGAATACTTTTGGCGTTCGGATCTTTATGCCAAGGTCTC 60
Db   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY   61 AGTCATAAGTTTTTRACCAGGAACAATGCACACTGCTGGAGCGTCCGGTTCAGAAAA 120
Db   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY   9430 TCTCTTTCTATTGGAGCTGGTGAAATTTATTCACGTCTCGCGGAGAGTGGTAGTGGGAG 9371
Db   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY   121 TCGACTTTGGCTCAACTGCTTGGCACACTTCACAAAACCAAGTTCGGACAGATCCCTTGTC 180
Db   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY   9370 ACCACGCTGCTAAATGTTATCGACACCTAGATTCAAAAGATAGTGGAAAGTATTAAT 9311
Db   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY   181 GAGGATGTAGACCTTCTGAACTCTCTACGCGTAAGCAACGGTTATACAGAHAAATACG 240
Db   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY   9310 AACGAGATGGAATATCAGA-----CGAAAAAGAGGTATGACTCTTAAAAAAGAGGTG 9257
Db   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY   241 GTGGGCTATTATTTCAAGATTATGCCCTTGATTCGCCGACAGGACAGTTAAATCAACCTT 300
Db   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY   9256 TTAGGTTTTATTTCCAANAATATCTATTGATGAAAATGAACAGHTTTAGAAAACCTTA 9197
Db   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY   301 CAGCTTGGCGTGGAAAAACACAAATGGCTGAAATTCCTCAAGTACTTTCATGCTGTGGT 360
Db   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY   9196 TCATTACAGGTGGGAAAAAATCGCAAGCTGATGATAGCATTTTGAAGAAGTAGGAATG 9137
Db   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY   361 CTGTAGTCGTTCGAGAAAAGCCAGTTTGAACCTCTCTGGTGGCGNACAAACGAACT 420
Db   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY   9136 GATGAAAGCTATTATTAGCAAAAAGATATACCAATTAAGCGGTGGAGAAAAAACACGGAT 9077
Db   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY   421 GGGTTGGCGCGGTACTGCTCAAAAATCCCCGAATAATTCGTGCTGATGAACCAACCGGA 480
Db   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY   9076 GCTATTGTGGCAATTTTACTCAAACCATTTTCAGTTATTACTTGGGACGAACCAACAGGC 9017
Db   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY   481 GCTCTAGANTTAACAAACAGTACGTAGTCTATAGAACATTTAGAGCACTGCGCCGACAA 540
Db   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY   9016 AACTTTGGATGATAAAACAAACACAAAATCATTTGAATTTATTTTAGCAATTAAGAAAAACAA 8957
Db   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY   541 GGCGCCACCGTTGTTGTTGCTACGCACTCGCCCTCTTCGGAATCAGCGGATACCAAT 600
Db   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY   8956 GTTAAACAAATNTTGTGTACATGACCCAGCAATATCTGCAAAAGCAGATCAATC 8897
Db   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY   601 AT 602
Db   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY   8896 AT 8895
Db   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 6
US-10-282-122A-27244
; Sequence 27244, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 27244
LENGTH: 1401
TYPE: DNA
ORGANISM: Mycoplasma genitalium
US-10-282-122A-27244

Query Match      12.2%; Score 74.6; DB 13; Length 1401;
Best Local Similarity 50.1%; Pred. No. 1.7e-13;
Matches 222; Conservative 0; Mismatches 209; Indels 12; Gaps 1;

QY   100 GGAGCGTCCGGTTCAGAAAAATCGACTTTGCTCAACTGCTGTGGCACACTTGACAAACCA 159
Db   805 GGCCTTTCAGGATCTGTGTAACCCACATTACTAAACATTATTTTCAGGGATGGATAGACT 864
QY   160 AGTTCCGGACAGATCCTTTGCGAGGATGTAGACCTTCTGAAACTCTCTACGCGTAAGCAA 219
Db   865 TCTAGTGGTAGTGTATTGTTCATGGTTATACATGATTGTTTAAATGATAGAAAGCTC 924
QY   220 CGGTTATACAGGAAAAATACGGTGGGCTATTTATTTTCAAGATTATGCTTTGATTTCCGAC 279
Db   925 ACTAAATTCGTCGTAAGATGTTGTTGTTACATCTTTTCAACAATACGGTTTATTACCTAAT 984
QY   280 AGGACAGCTTAAATTCACCTTCAGCTTGGGTGGGTAACAC-----AAATGG 327
Db   985 TTAACTTTAGAAAACATTTGAGTAGGAGCAATCTTCAACCAGATCCTAGTAAAGG 1044
QY   328 CCTGAAATTCCTCAAGTACTTCATGCTGTGTGTTGCTTTGAGTCGTTGAGGAAAAGCGAGTT 387
Db   1045 ATCAGCATTTGATGCACCTTTTAGAAGCGGTTGGATGGATGTTTGC AAAAGAGCTTCTCT 1104
QY   388 TTTCAGCTCTCTGGTGGGACAAACAGCACTGCGTTGGCCGGTACTGCTCAAAAT 447
Db   1105 AATGAATTTGATGGTGGGCAACAGCAACGTTTTCATTTGCAAGAGCTTTTCTCTAAAAAC 1164
QY   448 CCCCGAATAATCTGGCTGATGAACCAACCGGAGCTCTAGATTTTAAACAACAGTGAAGCTA 507
Db   1165 CCCTTATTATTTTGGTGTAGTAACCTACTTGGGGCACTTGTATCTTGAGATGATGCCAAAT 1224
QY   508 GTCATAGAGCATTTGAGAGCACT 530
Db   1225 GTTTTAAAAACAGTTTTTTAGCAAT 1247

RESULT 7
US-03-960-870-13
; Sequence 13, Application US/09960870
; Publication No. US20030134281A1
; GENERAL INFORMATION:
```


; APPLICANT: Evans, Glen
 ; TITLE OF INVENTION: NANOMACHINE COMPOSITIONS AND METHODS OF
 ; FILE REFERENCE: P-EA 4738
 ; CURRENT APPLICATION NUMBER: US/09/960,870
 ; CURRENT FILING DATE: 2001-09-20
 ; NUMBER OF SEQ ID NOS: 19
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 13
 ; LENGTH: 31241
 ; TYPE: DNA
 ; ORGANISM: M. genitalium
 ; US-09-960-870-13

Query Match 12.2%; Score 74.6; DB 10; Length 31241;
 Best Local Similarity 50.1%; Pred. No. 1e-12;
 Matches 222; Conservative 0; Mismatches 209; Indels 12; Gaps 1;
 100 GGAGCGTCCGTTTCAGGAAATCGACTTTCTCAACTGCTTGGCAGACCTTGACAAACCA 159
 8214 GGCCCTTCAGGATCTGGTAAACACCACTTACTAAACATTTATTCAGGGATGGATAGAGCT 8273
 160 AGTTCGGACAGATCCCTTGTTCGAGGATGTAGACCTTCTGAAACTCTCTACGCGTAAGCAA 219
 8274 TCTAGTGGTAGTGTATTGTCAATGGTTATTAACATGATTTGTTAAATGATAGAAAGCTC 8333
 220 CGGTTATACAGGAAATAATACGGTGGGCTATTATTTCAAGATTATGCTTGAATCCCGAC 279
 8334 ACTAAATTCGGTCAAAAGATGTTGGTTACATCTTTCAACAATACGGTTTATTACCTAAT 8393
 280 AGGACAGTTAAATTCACCTTCAGCTTCGGTGGGAAACAC-----AAATGG 327
 8394 TTAACAGTTAGAGAAACATTTAGATAGGAGCAAACTTCAACAGATCCTAGTAAAGG 8453
 328 CCTCAATTCCTCAAGTACTTTCATGCTGTTGGTCTTTCAGTGGTTCGAGGAAACCCAGTT 387
 8454 ATCAGCATTTGATGACATTTTGAAGCGTTGGGATGATAGTTTTCGAAAGAGAGCTTCCCT 8513
 388 TTTGAACCTCTCTGGTGGGAAACAAACAGAACTCGTTCGGTGGCCGGTACTGCTCAAAAT 447
 8514 AATGAATTTGAGTGGTGGCAACAGCAACGTTTTCATTCGAGAGCTTTGCTTAAACAC 8573
 448 CCCGAAATTCCTGGCTGATGAACCAACCGAGCTCTAGATTTAAACAAACAGTGAGCTA 507
 8574 CCCTATTATTTTGGTGAATGAACCTACTGGGCACTTGTATCTTGAGATGACCCAAAT 8633
 508 GTCATAGAAGCATTTGAGAGCACT 530
 8634 GTTTTAAACAGTTTTTAGCAAT 8656

RESULT 8
 US-09-960-858-13
 ; Sequence 13, Application US/09960858
 ; Publication No. US20030138777A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Evans, Glen
 ; TITLE OF INVENTION: NANOMACHINE COMPOSITIONS AND METHODS OF
 ; FILE REFERENCE: P-EA 4974
 ; CURRENT APPLICATION NUMBER: US/09/960,858
 ; CURRENT FILING DATE: 2001-09-20
 ; NUMBER OF SEQ ID NOS: 19
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 13
 ; LENGTH: 31241
 ; TYPE: DNA
 ; ORGANISM: M. genitalium
 ; US-09-960-858-13

Query Match 12.2%; Score 74.6; DB 10; Length 31241;
 Best Local Similarity 50.1%; Pred. No. 1e-12;
 Matches 222; Conservative 0; Mismatches 209; Indels 12; Gaps 1;

100 GGAGCGTCCGTTTCAGGAAATCGACTTTCTCAACTGCTTGGCAGACCTTGACAAACCA 159
 8214 GGCCCTTCAGGATCTGGTAAACACCACTTACTAAACATTTATTCAGGGATGGATAGAGCT 8273
 160 AGTTCGGACAGATCCCTTGTTCGAGGATGTAGACCTTCTGAAACTCTCTACGCGTAAGCAA 219
 8274 TCTAGTGGTAGTGTATTGTCAATGGTTATTAACATGATTTGTTAAATGATAGAAAGCTC 8333
 220 CGGTTATACAGGAAATAATACGGTGGGCTATTATTTCAAGATTATGCTTGAATCCCGAC 279
 8334 ACTAAATTCGGTCAAAAGATGTTGGTTACATCTTTCAACAATACGGTTTATTACCTAAT 8393
 280 AGGACAGTTAAATTCACCTTCAGCTTCGGTGGGAAACAC-----AAATGG 327
 8394 TTAACAGTTAGAGAAACATTTAGATAGGAGCAAACTTCAACAGATCCTAGTAAAGG 8453
 328 CCTCAATTCCTCAAGTACTTTCATGCTGTTGGTCTTTCAGTGGTTCGAGGAAACCCAGTT 387
 8454 ATCAGCATTTGATGACATTTTGAAGCGTTGGGATGATAGTTTTCGAAAGAGAGCTTCCCT 8513
 388 TTTGAACCTCTCTGGTGGGAAACAAACAGAACTCGTTCGGTGGCCGGTACTGCTCAAAAT 447
 8514 AATGAATTTGAGTGGTGGCAACAGCAACGTTTTCATTCGAGAGCTTTGCTTAAACAC 8573
 448 CCCGAAATTCCTGGCTGATGAACCAACCGAGCTCTAGATTTAAACAAACAGTGAGCTA 507
 8574 CCCTATTATTTTGGTGAATGAACCTACTGGGCACTTGTATCTTGAGATGACCCAAAT 8633
 508 GTCATAGAAGCATTTGAGAGCACT 530
 8634 GTTTTAAACAGTTTTTAGCAAT 8656

RESULT 9
 US-10-251-668-13
 ; Sequence 13, Application US/10251668
 ; Publication No. US20040063097A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Evans, Glen
 ; TITLE OF INVENTION: NANOMACHINE COMPOSITIONS AND METHODS OF
 ; FILE REFERENCE: P-EA 5441
 ; CURRENT APPLICATION NUMBER: US/10/251,668
 ; CURRENT FILING DATE: 2002-09-20
 ; PRIOR APPLICATION NUMBER: US 09/960,607
 ; PRIOR FILING DATE: 2001-09-20
 ; NUMBER OF SEQ ID NOS: 19
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 13
 ; LENGTH: 31241
 ; TYPE: DNA
 ; ORGANISM: M. genitalium
 ; US-10-251-668-13

Query Match 12.2%; Score 74.6; DB 13; Length 31241;
 Best Local Similarity 50.1%; Pred. No. 1e-12;
 Matches 222; Conservative 0; Mismatches 209; Indels 12; Gaps 1;
 100 GGAGCGTCCGTTTCAGGAAATCGACTTTGCTCAACTGCTTGGCAGACCTTGACAAACCA 159
 8214 GGCCCTTCAGGATCTGGTAAACACCACTTACTAAACATTTATTCAGGGATGGATAGAGCT 8273
 160 AGTTCGGACAGATCCCTTGTTCGAGGATGTAGACCTTCTGAAACTCTCTACGCGTAAGCAA 219
 8274 TCTAGTGGTAGTGTATTGTCAATGGTTATTAACATGATTTGTTAAATGATAGAAAGCTC 8333
 220 CGGTTATACAGGAAATAATACGGTGGGCTATTATTTCAAGATTATGCTTGAATCCCGAC 279
 8334 ACTAAATTCGGTCAAAAGATGTTGGTTACATCTTTCAACAATACGGTTTATTACCTAAT 8393
 280 AGGACAGTTAAATTCACCTTCAGCTTCGGTGGGAAACAC-----AAATGG 327

Db 8394 TTACAGTTAGAGAAACATTGAGATAGGAGCAAAATCTTCAACGAGATCTCTAGTAAAGG 8453
QY 328 CCTGAATTCCTCAAGTACTTCATGCTCTTGGTCTTGAGTCGTTGAGGAAACCCAGTT 387
Db 8454 ATCAGCATGTATGCACTTTTAGAGCGGTTGGGATGGATATTTGCAAAAGAGCTTCCT 8513
QY 388 TTGAACTCTCTGGTGGCGAAACAACAGCAACTGCTGGTGGCCGGTACTGCTCAAAAAT 447
Db 8514 AATGAATTCAGTGGTGGCAACAGCAAGCTGTTCCATTGCAAGAGCTTTTGCTAAAC 8573
QY 448 CCCGAATAATCTGGTGTAGTGAACCAACCGAGCTCTAGATTTAAACAACAGTGAGCTA 507
Db 8574 CCCATTATTAATTTTGGTGTAGTGAACCACTACTGGGCACTTGATCTTGAGATGACCAAAAT 8633
QY 508 GTCATAGAAGCATTTGAGAGCACT 530
Db 8634 GTTTTAAACAGTTTITAGCAAT 8656
RESULT 10
US-10-205-220-1
; Sequence 1, Application US/10205220
; Publication No. US20030170663A1
; GENERAL INFORMATION:
; APPLICANT: Fraser et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitium Genome, Fragment
; FILE REFERENCE: PB193P1D1
; CURRENT APPLICATION NUMBER: US/10/205,220
; PRIOR FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 08/545,528
; PRIOR FILING DATE: 1995-10-19
; PRIOR APPLICATION NUMBER: US 08/488,018
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/473,545
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 580073
; TYPE: DNA
; ORGANISM: Mycoplasma genitalium
US-10-205-220-1
Query Match 12.2%; Score 74.6; DB 15; Length 580073;
Best Local Similarity 50.1%; Pred. No. 5.4e-12;
Matches 222; Conservative 0; Mismatches 209; Indels 12; Gaps 1;
QY 100 GGAGCGTCCGGTTCAGGAAATCGACTTTGCTCAACTGTCTTGGCACACTTGACAAACCA 159
Db 78490 GGCCCTTCAGGATCTGTGTAACACACATTAACAATTTTCAGGGATGGATAGCT 78549
QY 160 AGTCCGAGAGATCTTGTTCGAGGATGATAGCTTCTGAACTCTCTACCGGTAAGCAA 219
Db 78550 TCTAGTGTATGTATGTGCAATGTGTTAATGATGATTTGTTTAAATGATAGAAAGCTC 78609
QY 220 CGGTATACAGGAAATAACGTTGGGCTATTATTCAAGATTATGCTTTGATTCGCGAC 279
Db 78610 ACTAAATTCGTCAAAAGATGTTGTTATATCTTCAACATACGTTTATACCTAT 78669
QY 280 AGGACATTAATCAACCTTCAGTTCGTTGGTGGGAAACAC-----AAATGG 327
Db 78670 TTAACGTTAGAGAAACATTGATAGAGAGCAAAATCTTCAACAGATCCTAGTAAAGG 78729
QY 328 CCTGAAATTCCTCAAGTACTTCATGCTTTGGTCTTTCAGTCTGTTTCAGGAAAGCCAGTT 387
Db 78730 ATCAGCATGTATGCACTTTTAGAGCGGTTGGATGATGATTTGCAAAAGAGCTTCCT 78789
QY 388 TTGAACTCTCTGGTGGCGAAACAACAGCACTGCGTTGGCCGGGTACTGCTCAAAAAT 447
Db 78790 AATGAATTCAGTGGTGGGCAACAGCAACGTTTCCATTGCAAGAGCTTTTGTCTAAAC 78849
QY 448 CCCGATAATTCCTGGCTGATGAACCAACCGAGCTCTAGATTTAAACAACAGTGAGCTA 507

Db 78850 CCCTTATTAATTTTGGTGTATGAACCTACTGGGCACCTGTATCTTGAGATGACCCAAAT 78909
QY 508 GTCATAGAAGCATTTGAGAGCACT 530
Db 78910 GTTTTAAACAGTTTITAGCAAT 78932
RESULT 11
US-10-205-220-1/c
; Sequence 1, Application US/10205220
; Publication No. US20030170663A1
; GENERAL INFORMATION:
; APPLICANT: Fraser et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitium Genome, Fragment
; FILE REFERENCE: PB193P1D1
; CURRENT APPLICATION NUMBER: US/10/205,220
; PRIOR FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 08/545,528
; PRIOR FILING DATE: 1995-10-19
; PRIOR APPLICATION NUMBER: US 08/488,018
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/473,545
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 580073
; TYPE: DNA
; ORGANISM: Mycoplasma genitalium
US-10-205-220-1
Query Match 12.2%; Score 74.6; DB 15; Length 580073;
Best Local Similarity 50.1%; Pred. No. 5.4e-12;
Matches 216; Conservative 0; Mismatches 209; Indels 6; Gaps 1;
QY 67 AAGTTTTTACCAGGAAACAATGACAGCACTGACTGAGCGTCCGGTTCAGGAAATCGACT 126
Db 576972 AAAATCTCTCCAGGAGAAATTTGTTTATTCTTGTAAATCTGTTAGTGTAAACCCAGC 576913
QY 127 TTGCTCAACTGTCTTGGCACACTTGACAAACCAAGTTCCGGACAGATCTTGTTCGAGAT 186
Db 576912 TTACTTTCTTTAAATTCGCAATTAGTCGACCACTAGCGCGATAGTTTGTGGA 576853
QY 187 GTAGACCTTCTGAACTCTCTACGCGTAAGCAACGTTTATACAGGAAATAACGTTGGC 246
Db 576852 ACTAACACCATATGTTGTAGTATGATGATGCAAACTAACCCGTTTAAAGAAATAAAACGTTGGT 576793
QY 247 TATTTATTTCAAGATTATGCTTTCATTCCTGATTCCTGCGACAGGACAGTTAAATTCACCTTCAGCTT 306
Db 576792 TATATTTTCAACATATGTTTATTTGCGTATGTTTACGCTTGTATGATATATTTAAGCTT 576733
QY 307 CGGTGGGAAACCAAAATGCGCTGAAATTCCTC-----AAGTACTTCTATGCTGTTGGT 360
Db 576732 GCTCTACCATTTAAAAAGGTTTAAATAATCTTGAAGAAATTTATTAGAAGCTTGGAA 576673
QY 361 CTTGAGTCGTTTCAGGAAACCAAGCTTTTGAAGCTCTCTGTCGCGAACAACAGCACT 420
Db 576672 TTAAGAAACATGACATAAGAAAGTTTCATAAACTCTCTGTTGGACAAACACGTTGTT 576613
QY 421 CGGTGGCCCGGCTACTGCTCAAAAATCCCGCAATAATTTCTGGCTGATGAACCAACCGGA 480
Db 576612 GCTATTCGCTGGCTTTAAATTAAGAACTTAAATTTTGTGTTGATGAACCTACTGGA 576553
QY 481 GCTCTAGATTT 491
Db 576552 GCAGTTAATAT 576542
RESULT 12
US-10-282-122A-34668
; Sequence 34668, Application US/10282122A

Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 34668
LENGTH: 756
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-10-282-122A-34668

Query Match 12.2%; Score 74.4; DB 13; Length 756;
Best Local Similarity 50.0%; Pred. No. 1.4e-13;
Matches 229; Conservative 0; Mismatches 211; Indels 18; Gaps 1;

Qy 57 TCTCAGTCATAAGTTTACAGAACAAATGACAGCACTGACCTGGAGCGTCCGGTTCAGG 116
Db 72 TATAAGTTTTCAGTTGAAAGGGGAATTTGTAGCTGTGTCATGGTCCCTTCTGGATCAGG 131

Qy 117 AAAATCGACTTTGCTCAACTGTCTTGGCACACTGTGACAAACCAAGTTTCGGACAGATCCT 176
Db 132 TAAACGACTTTACTCATGTGTTAAGTTCAATAGACTACTTTCAGGAGAACTGTAGA 191

Qy 177 TGTGAGAGTGTAGACCTCTGAACTCTCTACCGTAAAGCAAGGTTATACAGGAAATA 236
Db 192 AGTTGAAGGTAAAGAAATTAATAAATGAGCCAAAGAGTGGCAAAATTTTCGAAATA 251

Qy 237 TAAGGTGGCTTATTTATTTCAAGATTATGCTTATTCCTGATTCCTGACAGACAGTTAAATTC 296
Db 252 ACATCTCGGTTTATTTTTCAGATTATAGGTTTATACCATTAACAGTAAAGAAATA 311

Qy 297 CCTTCAGCTTGGGTGAAATAACAAATGGCTGAAATTCCTC----- 340
Db 312 TATTTATGCTACTCTCTCAGTACAAAATTCATAATATGAAATGGAACAAAATTTATAA 371

Qy 341 --AGTACTTCATCTCTGTCTTGTAGCTGTCGAGGAAAGCCAGCTTTTTCGAACTTC 398
Db 341 --AGTACTTCATCTCTGTCTTGTAGCTGTCGAGGAAAGCCAGCTTTTTCGAACTTC 398

Db 372 AGAGTGGCTGAGGCATTAGGTATTTATTAACCTGGGAAATAATATATCAAGTGAATTC 431
Qy 399 TGGTGGCGAACAAACAAACGAACTGCTTGGCCCGGCTACTGCTCAAAATCCCGCAATAAT 458
Db 432 TGGCGGTGAGCAACACGATACGGGGGAGCCCGGCGCATTCGTCATAAACCAACAATTAT 491
Qy 459 TCTGGCTGATGAACCAACCGGAGCTCTAGATTAAACAA 496
Db 492 TTTGCGAGATGAACCTACTGCGCATTAGATTCTAAAA 529

RESULT 13
US-10-398-221-3820
Sequence 3820, Application US/10398221
Publication No. US20040018514A1
GENERAL INFORMATION:
APPLICANT: KUNST, Frederik
APPLICANT: GLASER, Philippe
TITLE OF INVENTION: Listeria innocua, genome and applications
FILE REFERENCE: 344 702 - US
CURRENT APPLICATION NUMBER: US/10/398,221
CURRENT FILING DATE: 2003-03-27
PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
PRIOR FILING DATE: 2001-10-04
PRIOR APPLICATION NUMBER: FR 00/12 697
PRIOR FILING DATE: 2000-10-04
NUMBER OF SEQ ID NOS: 4025
SOFTWARE: PatentIn version 3.0
SEQ ID NO 3820
LENGTH: 5312
TYPE: DNA
ORGANISM: Listeria monocytogenes 4b
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(end)
OTHER INFORMATION: n can be any nucleotide: a,g,c or t/u
US-10-398-221-3820

Query Match 12.1%; Score 73.6; DB 16; Length 5312;
Best Local Similarity 50.2%; Pred. No. 7.9e-13;
Matches 234; Conservative 0; Mismatches 223; Indels 3; Gaps 2;

Qy 100 GGAGCGTCCGGTTCAGGAAATCGACTTTGTCTCAACTGTCTTGGCACACTTTCGACAAACA 159
Db 3955 GGACCTTCTGGAGCGGAAATCCACTGTACTTAATATACITGTGTGAATGATAGTCA 4014

Qy 150 AGTCCGGACAGATCCTTGTGAGATGTAGACCTTCTGAACCTCTCTACCGTAAGCAA 219
Db 4015 AGTGAAGGCAAAATATGCTAGATGGCAGGATATGCTCAATATAATGCAAAACAATA 4074

Qy 220 CGTTTATACAGAAATAATACGCTGGGCTATTTTATTTCAAGATTATGCTTGAATCCCGAC 279
Db 4075 ACCAATATCGCGNACAGATGTGGCTTTGTATTTCAATTTTACATTTGTACCCCAAT 4134

Qy 280 AGGACAGTTAAATTAACCTTCAGCTTGGCGGTGGAAA-AAACAAATGGCTGAAATTC 338
Db 4135 TTGACTGTCTAAAGAGACGTAGAAATTAGTCTGCCAAATTTGGCCAAATGCAATGACGCA 4194

Qy 339 TCAAGTACTTCA--TGCTGTTGGTCTTGGAGTGGTTCGAGGAAAGCCAGTTTTCGAACTC 396
Db 4195 GAACTGTTTAACTCAAGTTGGATTAAAGTACAGATTAGACAATTTTCAGACCACTA 4254

Qy 397 TCTGTGGCGAACCAACAGAACTGCGTTGGCCCGGTACTGCTCAAAAATCCCGAATA 456
Db 4255 TCTGGCGGGAACCAACAGCGTGTGCCATAGACGCTGGCTTGTCTTAAAGCACCAGAACTG 4314

Qy 457 ATTCTGGCTGATGAACCAACCGGAGCTCTAGATTAAACAACAGTACGCTAGCTATGAA 516
Db 4315 CTTTTATCGATGAGCCAAACCGGGCACTTGATTACGATCTGGAATAATCAGTCTTAAAA 4374

Qy 517 GCATTGAGAGCACTGCCGCAAAAGGCGCCACCGTTGTTGTTGCTA 562
Db 4375 TTATTACAGAACTTCCGTTAACAACGGGAACAACCGTTATTGTTA 4420

RESULT 14
US-10-282-122A-30817
; Sequence 30817, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent version 3.1
; SEQ ID NO 30817
; LENGTH: 684
; TYPE: DNA
; ORGANISM: Pasteurella multocida
US-10-282-122A-30817

Query Match 12.0%; Score 73.2; DB 13; Length 684;
Best Local Similarity 46.4%; Pred. No. 3.3e-13;
Matches 283; Conservative 0; Mismatches 318; Indels 9; Gaps 1;

QY 2 TCATAGAAATCAATGACCTCAAGAAATCTTTTGGCGTTGCGATCTTATGGCAAGGTCTCA 61
DB 2 TGATGAAACACACATTTATACAAAGATTCGTCAGTGACCGCACTTGAAGACATCA 61

QY 62 GTCATAAGTTTTTACCAGGAACAATGACAGCACTGACGTGGAGCGTCCGGTTTCAGGAAAT 121
DB 62 ACATTCAAATTAAGAGAGGCGCAATTTGTTGCAATCATGGTGCCTCAGGTTTCAGGCAAAA 121

QY 122 CGACTTTCCTCAACTGTCTTGGCACACTTGACAAACCAAGTTCGGGACAGATCTCTTGTG 181
DB 122 CGAGTTAATGAATTTCTACTGTTTAGCACTGCGATGAAGGAAAGTATCTCTAG 181

QY 182 AGGATGTAGACCTTCTGAAACTCTTACGGGTAAAGCAACGGTTATACAGGAAATAACGG 241
DB 182 ATGGTGTGATGCGGCACAACTTGAATTCGGCAGACACACGATTCGGGCGGAGAGA 241

QY 242 TGGGCTATTATTTCAGAGATTATGCCTTGATTCCTCCGACAGGACAGTTAAATTCACCTTC 301

DB 242 TCGGTCTGTCCTTTTCAGCAATTCACCTTAATCCCTTTATTTAACTGCCTTGAAACGTC 301
QY 302 AGCTTGGCGTGAACAAACACAAATGGCTGAAATTCCTCAA-----GTACTTTCATG 352
DB 302 TGTAGCACAAACATTTATCACAGTGAATTTGATGAAGATGCAGCCAAAGCGGTGTTTACAAC 361
QY 353 CTGTTGGTCTTCAGTCCGTTTCAGAGGAAAGCCAGTTTGTGAACTCTCTGCTGGCGAAACAAC 412
DB 362 AAGTCGGTCTGGCGCATCGTTTCGACACAGACCAAGCAACTGCTCGCGAGACAAC 421
QY 413 AAGCAACTGCGTTGGCCCGGTACTGCTCAAAAATCCCGAATTAATTCGCTGATGAC 472
DB 422 AAGCTGTGTGTTATTCGCGCGCTTTAGTTAATCAGCCCGCGTGATTTTCCCGATGAGC 481
QY 473 CAACCGGAGCTCTAGATTTAACAAACAGTGTAGTCTATAGAAGCATTTGAGAGCACTCG 532
DB 482 CGACAGGAAACCTAGATGAAAAAATGAGCACTGCTCGATTACTCGTCACATTAA 541
QY 533 CGACAGGAAAGCGCCACCGTTGTTGTTGCTAGCACTCGCCCTCTTCGAGAAATCAGCGG 592
DB 542 ACCAAACAGAGCGTACAGTGTGATGTCACCCCAATATCTGAGCTAAGTAACTGGCGC 601
QY 593 ATACCATTAAT 602
DB 602 ACCGCACTAT 611

RESULT 15
US-09-738-626-956
; Sequence 956, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patent ver. 3.0
; SEQ ID NO 956
; LENGTH: 666
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-956

Query Match 11.9%; Score 72.4; DB 9; Length 666;
Best Local Similarity 47.4%; Pred. No. 6e-13;
Matches 257; Conservative 0; Mismatches 276; Indels 9; Gaps 1;

QY 76 CCAGGAAACAATGACAGCACTGACTGGAGCGTCCGGTTTCAGGAAATCGACTTTGCTCAAC 135
DB 106 CCGCGGAACCTCGTGGCCATCATGGCCGCTCCGGCTCAGGAAATCCACCTTCTCAAC 165

QY 136 TGCTTTGGCACATTTGACAAACCAAGTTCCGGACAGATTCCTTGTGAGGATGTAGACCTT 195
DB 166 GTCCGCGGCTCTCTGACGGGCAACCTCTCGCCATGTGCTTATCGAGGTGCCAGCGCC 225

QY	196	CTGAACCTCTCTACGGTAAACACGGTTATACAGGAAAAATACGGTGGCTATTATT	255
Db	226	TCAGACCTCAACGCCAAACGGCGAGCTGAACACCGCGCTCGCCACATCGAGTTATTTTC	285
QY	256	CAAGATTATGCTTGTATCCCGACAGACAGATTAAATTCACCTTCAGCTTCGGTGGAA	315
Db	286	CAAACTACAACTGTGTCCTCCACCTCACCGTCGGAGAAAACGTGGTCTGCCCTAGAA	345
QY	316	-----AAACACAAATGGCTGAAATTCCTCAAGTACTTCATGCTGTTGTTGAG	366
Db	346	CTCGACGGCAAAACCGACCGCCAGGAGTAGCAATCGCACTCGGGAAGTCGGCCTCGAA	405
QY	367	TCGTTTCGAGGAAAAGCCAGTTTTTGAACCTCTCTGGTGGCGAAACACACGAACTGCGTTG	426
Db	406	GGCTTCGACGACCGCTTCCCGGAGAGATCTCGGCGGCCAAGCCAGCGCTCGCCATC	465
QY	427	GCCCGGTACTGCTCAAAATCCCGAATAATTCGCTGATGAACCAACCGAGCTCTA	486
Db	466	GCGAGAGCCCTCATCGGCCCGCGAAAAATCTTGCTTCCGACGAAACCCACCGCGCCCTC	525
QY	487	GATTTAACRAACAGTGAAGTGTAGTATAGAGCAATTCAGAGCACTCGCCGACAAAGCGGC	546
Db	526	GACACCTCCACCGGACGCGAGTCTCCGGCTCTCCGCCAAGAATCGATTCGGTGCC	585
QY	547	ACCGTTGTTGTTGCTACGCACTCGCCCTCTTCCGAGATTCAGCGGATACCATTTATCAA	606
Db	586	GCAGGCTCTTGTCAACCAAGAACCCCGTTCCGCGGTGGGCGAGACCGAAACAAATCATG	645
QY	607	CT 608	
Db	646	CT 647	

Search completed: July 30, 2004, 15:28:19
Job time : 518 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 30, 2004, 07:32:14 ; Search time 2796 Seconds
(without alignments)
6504.316 Million cell updates/sec

Title: US-09-868-338-7_COPY_1117_1725

Perfect score: 609

Sequence: 1 atgatgaaatcaatgacct.....cggtaccattatcaacta 609

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

1: em_estba.*

2: em_esthum.*

3: em_estin.*

4: em_estma.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_hc.*

9: gb_est1.*

10: gb_est2.*

11: gb_hc.*

12: gb_est3.*

13: gb_est4.*

14: gb_est5.*

15: em_estfun.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inv.*

19: em_gss_pin.*

20: em_gss_vit.*

21: em_gss_fun.*

22: em_gss_man.*

23: em_gss_mus.*

24: em_gss_pro.*

25: em_gss_rod.*

26: em_gss_phg.*

27: em_gss_vrl.*

28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	75.2	12.3	647	14	CD164788
2	63.4	10.4	622	28	BH861070
3	62.2	10.2	639	29	CG954936
4	62.2	10.2	668	14	CA847456

5	61	10.0	404	14	CD740366
6	61	10.0	542	14	CD731492
7	61	10.0	619	9	AL586817
8	61	10.0	743	13	BU399305
9	61	10.0	798	13	BU328328
10	61	10.0	896	13	BU407162
11	61	10.0	907	13	BU112595
12	60.4	9.9	759	13	BU865169
13	59.4	9.8	576	9	AU090519
14	59.2	9.7	358	14	H54613
15	59.2	9.7	426	10	BF700602
16	59.2	9.7	448	10	BF691964
17	59.2	9.7	481	13	EX281688
18	59.2	9.7	528	9	AA456377
19	59.2	9.7	542	12	BM126834
20	59.2	9.7	584	12	BT195085
21	59.2	9.7	598	9	AT807162
22	59.2	9.7	646	14	CD672146
23	59.2	9.7	667	13	EX509020
24	59.2	9.7	761	13	EX106160
25	59.2	9.7	919	13	EX403383
26	59.2	9.7	923	13	BQ229972
27	59.2	9.7	963	12	BG387740
28	59.2	9.7	1201	9	AL520322
29	58.4	9.6	296	10	AY125863
30	57.8	9.5	713	9	AU213416
31	57.8	9.5	922	12	BG442902
32	57.6	9.5	780	9	AV726232
33	57.6	9.5	880	13	BX392299
34	57.6	9.5	1065	13	BX325244
35	57.6	9.5	4035	11	EC042531
36	57.4	9.4	790	14	CB894287
37	57.2	9.4	300	28	AZ578156
38	57	9.4	833	14	CB894924
39	56.8	9.3	515	28	BZ331622
40	56.8	9.3	566	9	AU165128
41	56.8	9.3	655	28	BZ348974
42	56.8	9.3	1201	9	AL520321
43	56.4	9.3	3843	29	AY408954
44	56.4	9.3	3843	29	AY408955
45	55.4	9.1	780	12	BG587938

ALIGNMENTS

RESULT 1
CD164788
LOCUS
DEFINITION
MLJ-0091T-L258-D10-U.B MLJ-0091 Schistosoma mansoni CDNA Clone
MLJ-0091T-L258-D10.B, mRNA sequence.
ACCESSION
CD164788
VERSION
EST.
KEYWORDS
SOURCE
ORGANISM
Schistosoma mansoni
Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
Strigeidida; Schistosomatidae; Schistosomatidae; Schistosoma.
REFERENCE
1 (bases 1 to 847)
Verjovskii-Almeida, S., DeMarco, R., Martins, E.A.L., Guimaraes, P.E.M.,
Ojopi, E.P.B., Paquola, A.C.M., Piazza, J.P., Nishiyama, M.Y. Jr.,
Kitajima, J.P., Adamson, R.E., Ashton, P.D., Bonaldo, M.P.,
Coulson, P.S., Dillon, G.P., Farias, L.P., Gregorio, S.P., Ho, P.L.,
Leite, R.A., Malaquias, L.C.C., Marques, R.C.P., Miyasato, P.A.,
Nascimento, A.L.T.O., Ohlweiler, F.P., Reis, E.M., Ribeiro, M.A.,
Sa, R.G., Stukart, G.C., Soares, M.B., Gargioni, C., Kawano, T.,
Rodrigues, V., Madeira, A.M.B.N., Wilson, R.A., Menck, C.F.M.,
Setubal, J.C., Leite, L.C.C. and Dias-Neto, E.
Transcriptome analysis of the acelomate human parasite Schistosoma mansoni

Nat. Genet. 35 (2), 148-157 (2003)
JOURNAL
22879226
MEDLINE
COMMENT
Contact: Dr. Sergio Verjovskii-Almeida

Departamento de Bioquímica
 Instituto de Química - Universidade de São Paulo
 Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 São Paulo - SP,
 Brasil
 Tel.: +55-11-3091-2173
 Fax: +55-11-3091-2186
 Email: verjoeiq.usp.br
 This sequence was derived from the FAPESP Schistosoma mansoni EST
 Genome Project. All sequences in the project were assembled and
 annotated. This entry and all the assembled sequences can be seen
 in the following URL <http://bioinfo.iq.usp.br/schisto/>
 Plate: ML1-00917-L258 row: 10 column: D.

FEATURES

Location/Qualifiers
 1..647
 /organism="Schistosoma mansoni"
 /mol_type="mRNA"
 /db_xref="taxon:6183"
 /clone="ML1-00917-L258-D10.B"
 /sex="mixed pool"
 /dev_stage="miracidium"
 /clone_lib="ML1-0091"
 /note="Vector: pGEM T-easy"

ORIGIN

Query Match 12.3%; Score 75.2; DB 14; Length 647;
 Best Local Similarity 53.8%; Pred. No. 2.4e-12;
 Matches 155; Conservative 0; Mismatches 133; Indels 0; Gaps 0;
 QY 1 ATGATAGAAATCAATGACCTCAAGAAATCTTTGGCGTTCGGATCTTATGGCAAGGCTC 60
 Db 222 ATGCTCGAAATCTCGAGCTGGCCAAATCGTTGGCTCCAGAGCGTTGTGGAGGGCCG 291
 QY 61 AGTCATAGTTTTCACAGGAACAAATGACAGCAGTCTGAGCGTCCGGTTCAGGAAA 120
 Db 292 AACTGGACGTTTCATGGCGGAAGGATGACACCGTGTGGGTTCGAGTGGCTGGGAAA 351
 QY 121 TGGATTGCTCACTGTCITGGCAGACTTGACAAACCAAGTTCGGACAGATCCTTGTC 180
 Db 352 TCGACACTGTTGAATGCTGCTGGAAACATGTGATCGTCCAGCGGTGAGCGTATCCTTTGG 411
 QY 181 GAGGATGTAGACCTTCTGAAACTCTCTACGGTGAAGCAACGGTTATACAGGAAAAATACG 240
 Db 412 AATGTAAGATGTGCTGCTGCAAGCGTCTGTCAGAGGCGACGATGCGGAAATCTGAC 471
 QY 241 GTGGCTATTTATTTCAAGATATGCTTGATTCGGACAGGACAGTT 288
 Db 472 CTCGGTTATTTATTCAGAACTACGCATTTGTTGAGCAGCCGCCGTT 519

RESULT 2

BH61070 622 bp DNA linear GSS 29-MAY-2003
 S2H2SP6 Western x-disease phytoplasma SpeI cosmid library Western x
 phytoplasma genomic clone S2H2 similar to ABC transporter, ATP
 binding protein (pHnL), genomic survey sequence.

ACCESSION

VERSION BH61070.1 GI:31126928

KEYWORDS

SOURCE Western X phytoplasma

ORGANISM

Western X phytoplasma
 Bacteria; Firmicutes; Mollicutes; Achleplasmatales;
 Achleplasmataceae; Phytoplasma; 16SIII (X-disease group).

REFERENCE

1. (bases 1 to 622)
 Liefting L.W. and Kirkpatrick B.C.
 Cosmid cloning and sample sequencing of the genome of the
 uncultivable mollicute, Western X-disease phytoplasma, using DNA
 purified by pulsed-field gel electrophoresis
 FEMS Microbiol. Lett. 221 (2), 203-211 (2003)

JOURNAL

MEDLINE 22612896

PUBMED

12725928

COMMENT

Contact: Liefting LW
 Department of Plant Pathology
 University of California, Davis

One Shields Ave, Davis, CA 95616, USA

Tel: 530 752 1697

Fax: 530 752 5674

Email: cwlietfing@ucdavis.edu

Sequences were manually edited using Sequencher to remove vector
 and low quality sequence.

Seq primer: SP6

Class: cosmid ends.

FEATURES

Location/Qualifiers

1..622

/organism="Western X phytoplasma"

/mol_type="genomic DNA"

/strain="Jensen"

/db_xref="taxon:37704"

/clone_lib="Western X-disease phytoplasma SpeI cosmid
 library"

/note="Vector: pcosRW2; Site 1: SpeI; Pulsed-field gel
 purified DNA from Western X-disease phytoplasma was
 digested to completion with SpeI and cloned into the AvrII
 site of pcosRW2. pcosRW2 was first modified by inserting
 an annealed oligonucleotide containing the recognition
 sequences for AvrII, NcoI and MluI into the EcoRI site."

ORIGIN

Query Match 10.4%; Score 63.4; DB 28; Length 622;
 Best Local Similarity 48.5%; Pred. No. 1.4e-08;
 Matches 215; Conservative 0; Mismatches 216; Indels 12; Gaps 1;
 QY 100 GGAGCGTCCGGTTCAGGAAATCGACTTTGCTCAACTGCTTTGGCACACTTGACAAACCA 159
 Db 156 GGAATATCAGGATCGGCAATCGACTCTTTAACTTTGGGAGGACTCGATAAGTGT 215
 QY 160 AGTTCGGACAGATCCTTGTGAGGATGACACTCTTGAACCTCTACCGTAAGCAA 219
 Db 216 GATAGTGGAGATATATTAAATTAAGTTTCTATAAAGAAATTAACAACAAATGATTA 275
 QY 220 CGGTTATACAGAAAAATACGGTGGCTATTATTTCAGAGATTATGCTTGAATCCCGAC 279
 Db 276 GACGACTATCGAAGCGGCTTTAGCTTTATTTTCCAAGAAATTAATTTAATAGAAAT 335
 QY 280 AGGACAGTTAAATTCACACT-----TCAGTTTCGGTGGAAAAACACAAATGG 327
 Db 336 ATGAATGTTTATGAAATATCGCTTTAGCTTATGAGTTTCAGGGCCAAAAACCGGACAT 395
 QY 328 CCTGAAATTCCTCAAGTACTTCATCTGTTGCTTGTAGTCTTCGAGGAAAAACGCACTT 387
 Db 396 AAACATATTATTAACCTTATTAAGAAATGGAACCTTTAGTTATGAAGAAAAAAT 455
 QY 388 TTTGAACCTCTGTTGGGCAACAAACGAACTGGCTTGGCCCGGTTACTGCTCAAAAT 447
 Db 456 AACGAATATCAGCGGCAAAAAACAAGAAATAGCCATCGCTCGTCTTTGGTTAAAAAT 515
 QY 448 CCCCAGATTAATTCGCTGATGAACCAACCGAGCTCTAGATTTTAAACAAACAGTGAGCTA 507
 Db 516 CCGGAGATTATTTTACGGATGAACCAACAGGAACTTGATTCAAAACCTGGCGAACAA 575
 QY 508 GTCATAGAGCACTTGAGAGCACT 530
 Db 576 ATTTTTAAATTTTAAAAAAAT 598

RESULT 3

CG954936

LOCUS CG954936

DEFINITION MBEP107F mth2 Medicago truncatula genomic clone 20B20, genomic
 survey sequence.

ACCESSION CG954936

VERSION CG954936.1

KEYWORDS GSS.

SOURCE Medicago truncatula (barrel medic)

ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

CG954936 639 bp DNA linear GSS 15-DEC-2003

MBEP107F mth2 Medicago truncatula genomic clone 20B20, genomic
 survey sequence.

CG954936

CG954936.1 GI:39870005

GSS.

Medicago truncatula (barrel medic)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 639)
Town, C.D., Shetty, J., Koo, H. and Feldblyum, T.P.
Sequencing of BAC ends from Medicago truncatula
Unpublished (2003)
Other GSSs: MBEBP10TR
Contact: Chris Town

TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

Seq primer: TGTAAACGACGGCCAGT

Class: BAC ends.

FEATURES
source

Location/Qualifiers

1..639

/organism="Medicago truncatula"

/mol_type="genomic DNA"

/cultivar="genotype Al7"

/db_xref="taxon:3880"

/clone="20B20"

/clone_lib="mth2"

/note="Vector: pBeloBAC11; Site 1: HindIII; Site 2:

HindIII; Cook, D.R. and Kim, D.J., unpublished"

ORIGIN

Query Match 10.2%; Score 62.2; DB 29; Length 639;
Best Local Similarity 56.7%; Pred. No. 3.3e-08;
Matches 115; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 358 GGTCTTGAGTCGTCGAGGAAAGCCAGTTTGAACCTCTCTGTCGGCGAACAACGA 417

DB 87 GGTCTTGATTCATGTGTAGGAGTCATGGGACTCAACTTCTCTGTCGGCAAAAGCAAGA 146

QY 418 ACTGCGTGTGCGCGGTACTGCTCAAAATCCCGAATAATCTGCTGATGAACCAACC 477

DB 147 ATTGCATTCGAGGGCCATTGAGATCCAGATCCTCTTCTTGATGAAGCAACA 206

QY 478 GGAGCTCTAGATTAAACAACAGTAGCTAGTCATAGAACATGAGAGCACTGCGCGAC 537

DB 207 AGTGCAATGTGATGCTGAGTCTGAACGTGTGTTCAAGAGCACTGGAAGAAAGTTATGACA 566

QY 538 AAAGGCGCCACCGTTGTTGTC 560

DB 267 CAAGAGCACTACCGTAGTTGTC 289

RESULT 4

CA847456

LOCUS

DEFINITION EST0364 CM334 Root cDNA Capsicum annuum cDNA clone RR29-81 similar to Arabidopsis thaliana putative ABC transporter protein, mRNA

Sequences.

ACCESSION CA847456

VERSION CA847456.1

KEYWORDS EST.

SOURCE Capsicum annuum

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

asterids; lamids; Solanales; Solanaceae; Capsicum.

1 (bases 1 to 668)

Richins, R., Alvarado, K., Leger, J. and O'Connell, M.A.

Responsive transcripts in Phytophthora capsici-challenged roots of

Capsicum annuum

Unpublished (2002)

Contact: O'Connell, MA

Department of Agronomy and Horticulture

New Mexico State University

MSC 30, P.O. Box 30003, Las Cruces, NM 88003, USA

Tel: 505 646 5172

Fax: 505 646 6041

Email: mococonnell@nmsu.edu

Seq primer: T3

High quality sequence stop: 668.

FEATURES

source

Location/Qualifiers

1..668

/organism="Capsicum annuum"

/mol_type="mRNA"

/strain="Criollo de Morelos 334 (CM334)"

/db_xref="taxon:4072"

/clone="RR29-81"

/dev_stage="6 hours after Phytophthora capsici

inoculation"

/clone_lib="CM334 Root cDNA"

/note="Organ: Root; Vector: Zap; Site_1: EcoRI; Site_2:

Xho 1"

ORIGIN

Query Match 10.2%; Score 62.2; DB 14; Length 668;

Best Local Similarity 56.7%; Pred. No. 3.4e-08;

Matches 115; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 358 GGTCTTGAGTCGTCGAGGAAAGCCAGTTTGAACCTCTCTGTCGGCGAACAACGA 417

DB 162 GGTCTAGACACATGGTTGGAGACATGGGACTCACTTCTGTCGGCAAAAGCAAGA 221

QY 418 ACTGCGTGTGCGCGGTACTGCTCAAAATCCCGAATAATCTGCTGATGAACCAACC 477

DB 222 ATTGCCATCGCAAGAGCAATTTAAAGATCCACGGATATTACTGTTAGATGAAGCAACA 281

QY 478 GGAGCTCTAGATTAAACAACAGTGAGCTAGTCATAGAACGATGAGAGCACTGCGCGAC 537

DB 282 AGTGATTAGATGAGATCTGAGAGGTGTACAGAGCATTTGGATAGATTATGATC 341

QY 538 AAAGCGCCACCGTTGTTGTC 560

DB 342 AACAGAACCAACATCGTAGTTGC 364

RESULT 5

CD740366

LOCUS

DEFINITION

CD740366

VERSION

CD740366.1

KEYWORDS

EST.

SOURCE

ORGANISM

Gallus gallus (chicken)

Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Archosauria; Aves; Neognathae; Galliformes; Phasianidae;

Phasianinae; Gallus.

1 (bases 1 to 404).

Tassel, C. and Han, J.Y.

Chicken intestinal lymphocyte EST database as a resource for the

analysis of mucosal immune function

Unpublished (2003)

Contact: Hyun S. Lillehoj

Animal Parasite Diseases Laboratory

Animal and Natural Resources Institute, USDA

Bldg. 1043, BARC-East, Beltsville, MD 20705, USA

Tel: 3015048771

Fax: 3015045103

Email: hlillehoenri@ars.usda.gov

Single pass sequencing. Bases called and trimmed with phred

0.000925 using options -trim_alt -trim_fasta. Vector identified

by cross_match using options -minmatch 12 -minscore 12

Plate: 24 row: H column: 07

Seq primer: ATTAGTGACACTATAG

High quality sequence stop: 404.

FEATURES

source

Location/Qualifiers

1..404

QY 358 GGTCTTGAGTCGTTGAGGAAAAGCCAGTTTTTTGAACCTCTCTGGTGGGAAACAAACGA 417
 |||||
 Db 176 GGAATTTGACACTGTAGTTGGAGAAAAGGCAATTTCTATTTCAGGTGGACAGAGCAGCA 235
 |||||
 QY 418 ACTGCGTTGGCCGGGTACTGCTCAAAAATCCCCGAAATTAATCTGCTGATGAACCAACC 477
 |||||
 Db 236 ATTGCAATTGCTGAGCTCTGCTCAGAATCCCAAAATTTCTTGTAGACGAGCAACA 295
 |||||
 QY 478 GGAGCTCTAGATTATCAAAACAGTGAAGCTAGTCATAGAAGCATTTG 522
 |||||
 Db 296 AGTGCTCTGGATGCTGAATAATGAGTACCTAGTCAAGAAGCTCTG 340
 |||||

RESULT 8
 BUI39305
 LOCUS 603535206F1 CSEQHN58 743 bp mRNA linear EST 27-NOV-2002
 DEFINITION 603535206F1 CSEQHN58 Gallus gallus cdna clone CHEST49511 5', mRNA
 sequence.

ACCESSION BUI39305
 VERSION BUI39305.1 GI:25768361
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus

REFERENCE 1 (bases 1 to 743)
 AUTHORS Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
 Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
 TITLE A Comprehensive Collection of Chicken cDNAs
 JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
 MEDLINE 22335534
 PUBMED 12445392

COMMENT Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology
 (UMIST)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.
 Location/Qualifiers

FEATURES
 source
 1..743
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="Compton Line 151"
 /db_xref="taxon:9031"
 /clone="CHEST49511"
 /sex="female"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_lib="CSEQHN58"
 /notes="Organ: small intestine; Vector: pBluescript II
 KS(+); Site 1: EcoRI; Site 2: NotI; This normalized
 library was constructed from 1 million independent clones.
 cDNA synthesis was initiated using an oligo(dT) primer,
 using methylated C in the first strand synthesis reaction.
 Following this first strand reaction, double-stranded cDNA
 was blunted, ligated to NotI adapters, digested with
 EcoRI, size-selected, and cloned into the NotI and EcoRI
 compatible sites of a custom modified MCS of the
 pBluescript (KS+) vector. The library was normalized in 2
 rounds using conditions adapted from Soares et al., PNAS
 (1994) 91: 9228-9232 and Bernaldo et al., Genome Research 6
 (1996): 791, except that a significantly longer
 reannealing hybridization was used."

ORIGIN

Query Match 10.0%; Score 61; DB 13; Length 743;
 Best Local Similarity 60.6%; Pred. No. 8.6e-08;
 Matches 100; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 358 GGTCTTGAGTCGTTGAGGAAAAGCCAGTTTTTTGAACCTCTCTGGTGGGAAACAAACGA 417
 |||||
 Db 162 GGAATTTGACACTGTAGTTGGAGAAAAGGCAATTTCTATTTCAGGTGGACAGAGCAGCA 221
 |||||
 QY 418 ACTGCGTTGGCCGGGTACTGCTCAAAAATCCCCGAAATTAATCTGCTGATGAACCAACC 477
 |||||
 Db 222 ATTGCAATTGCTGAGCTCTGCTCAGAATCCCAAAATTTCTTGTAGACGAGCAACA 281
 |||||
 QY 478 GGAGCTCTAGATTATTAACAAACAGTGAAGCTAGTCATAGAAGCATTTG 522
 |||||
 Db 282 AGTGCTCTGGATGCTGAATAATGAGTACCTAGTCAAGAAGCTCTG 326
 |||||

RESULT 9

BUI328328
 LOCUS 603496289F1 CSEQHN64 798 bp mRNA linear EST 28-NOV-2002
 DEFINITION 603496289F1 CSEQHN64 Gallus gallus cdna clone CHEST40624 5', mRNA
 sequence.

ACCESSION BUI328328
 VERSION BUI328328.1 GI:25836329
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus

REFERENCE 1 (bases 1 to 798)
 AUTHORS Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
 Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
 TITLE A Comprehensive Collection of Chicken cDNAs
 JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
 MEDLINE 22335534
 PUBMED 12445392

COMMENT Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology
 (UMIST)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.
 Location/Qualifiers

FEATURES
 source
 1..798

/organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="White Leghorn, HiseX"
 /db_xref="taxon:9031"
 /clone="CHEST40624"
 /tissue_type="whole embryo"
 /dev_stage="10"
 /lab_host="DH10B"
 /clone_lib="CSEQHN64"
 /notes="Organ: whole embryo; Vector: pBluescript II KS(+);
 Site 1: EcoRI; Site 2: NotI; This normalized library was
 constructed from 1 million independent clones. cDNA
 synthesis was initiated using an oligo(dT) primer, using
 methylated C in the first strand synthesis reaction.
 Following this first strand reaction, double-stranded cDNA
 was blunted, ligated to NotI adapters, digested with
 EcoRI, size-selected, and cloned into the NotI and EcoRI
 compatible sites of a custom modified MCS of the
 pBluescript (KS+) vector. The library was normalized in 2
 rounds using conditions adapted from Soares et al., PNAS
 (1994) 91: 9228-9232 and Bernaldo et al., Genome Research 6
 (1996): 791, except that a significantly longer
 reannealing hybridization was used."

ORIGIN

Query Match 10.0%; Score 61; DB 13; Length 798;
 Best Local Similarity 54.1%; Pred. No. 8.9e-08;
 Matches 124; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

QY 358 GGTCTTGAGTCGTTGAGGAAAAGCCAGTTTTTTGAACCTCTCTGGTGGGAAACAAACGA 417

Db 333 GGATTGACACTGTAGTTGGAGAAAAGGCACTTCTCTTTTCAGGTGGACAGAGCAGCGA 392
 QY 418 ACTCGTTGGCCGGGTACTGCTCAAAAATCCCGGAATAATTTCTGGCTGATGAACCAACC 477
 Db 393 ATTGCATTTCTCAGCTCTGCTCAAGATCCCAAAATTTCTTTGTTAGACGAGCAACA 452
 QY 478 GGACTCTAGATTAAACAAAGTACGTAGTATAGAACATTTAGAGACATTCGCGGCAC 537
 Db 453 AGTCTCTGATGCTGAAATAGTACCTAGTCAAGAGCTCTGACACCGCTCATGGCA 512
 QY 538 AAGGGCCCGCTGTTGTTGTACCACTCGCCCTCTTCGAGAAAT 586
 Db 513 AGGAGGACAGCTCTTAATCATCTGCTACCGTCTGCTTACTATTTCAGAAAT 561

RESULT 10
 BU407162
 LOCUS 896 bp mRNA linear EST 27-NOV-2002
 DEFINITION BU407162 896 bp mRNA linear EST 27-NOV-2002
 sequence.
 ACCESSION BU407162
 VERSION BU407162
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 1 (bases 1 to 896)
 Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
 A Comprehensive Collection of Chicken CDNAs
 Curr. Biol. 12 (22), 1965-1969 (2002)
 2235534
 MEDLINE
 PUBMED 12445392
 COMMENT Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology
 (UMIST)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.
 Location/Qualifiers

FEATURES
 source
 1..896
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="White Leghorn, Hisex"
 /db_xref="taxon:9031"
 /clone="CHEST375f3"
 /dev_stage="36"
 /lab_host="DH10B"
 /clone_lib="CSEQCHN59"
 /note="Organ: limbs; Vector: pBluescript II KS(+); Site_1:
 EcoRI; Site_2: NotI; This normalized library was
 constructed from 1 million independent clones. cDNA
 synthesis was initiated using an oligo(dT) primer, using
 methylated C in the first strand synthesis reaction.
 Following this first strand reaction, double-stranded cDNA
 was blunted, ligated to NotI adapters, digested with
 EcoRI, size-selected, and cloned into the NotI and EcoRI
 compatible sites of a custom modified MCS of the
 pBluescript (KS+) vector. The library was normalized in 2
 rounds using conditions adapted from Soares et al., PNAS
 (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
 (1996): 791, except that a significantly longer
 reannealing hybridization was used."

ORIGIN

Query Match 10.0%; Score 61; DB 13; Length 896;
 Best Local Similarity 60.6%; Pred. No. 9.4e-08;
 Matches 100; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 358 GGTCTTGTAGTCTGTTCGAGAAAAGCCAGTTTGTGAACCTCTCTGTGGCGAACAACACGA 417
 Db 162 GGATTGACACTGTAGTTGGAGAAAAGGCACTTCTCTTTTCAGGTGGACAGAGCAGCGA 221
 QY 418 ACTCGTTGGCCGGGTACTGCTCAAAAATCCCGGAATAATTTCTGGCTGATGAACCAACC 477
 Db 222 ATTGCATTTCTCAGCTCTGCTCAAGATCCCAAAATTTCTTTGTTAGACGAGCAACA 281
 QY 478 GGAGCTCTAGATTAAACAAAGTACGTAGTATAGAACATTTAGAGACATTCG 522
 Db 282 AGTCTCTGATGCTGAAATAGTACCTAGTCAAGAGAAAGCTCTG 326

RESULT 11
 BU112595
 LOCUS 907 bp mRNA linear EST 25-NOV-2002
 DEFINITION BU112595 907 bp mRNA linear EST 25-NOV-2002
 sequence.
 ACCESSION BU112595
 VERSION BU112595.1 GI:25316826
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 1 (bases 1 to 907)
 Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
 A Comprehensive Collection of Chicken CDNAs
 Curr. Biol. 12 (22), 1965-1969 (2002)
 2235534
 MEDLINE
 PUBMED 12445392
 COMMENT Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology
 (UMIST)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.
 Location/Qualifiers

FEATURES
 source
 1..907
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="White Leghorn, Hisex"
 /db_xref="taxon:9031"
 /clone="CHEST102n17"
 /dev_stage="22"
 /lab_host="DH10B"
 /clone_lib="CSEQCHL14"
 /note="Organ: head; Vector: pBluescript II KS(+); Site_1:
 EcoRI; Site_2: NotI; Modification of pBluescript II KS(+) with
 [Stratagene] vector to accommodate cDNA produced with the
 T-stranded protocol (Construction of uni-directionally
 cloned cDNA libraries from messenger RNA for improved 3'
 end DNA sequencing by Glenn Fu, et al. U.S. Patent #
 6,387,624). Cut pBluescript II KS(+) with NotI and EcoRI.
 Ligate in double stranded adaptor containing BsgI and
 BamHI sites [5'ggcggctgcagccgcgcatccgaaaaag
 [5'aattcttttttcggatccgggctgcagc]]

ORIGIN

Query Match 10.0%; Score 61; DB 13; Length 907;
 Best Local Similarity 60.6%; Pred. No. 9.4e-08;
 Matches 100; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 358 GGTCTTGTAGTCTGTTCGAGAAAAGCCAGTTTGTGAACCTCTCTGTGGCGAACAACACGA 417
 Db 394 GGATTGACACTGTAGTTGGAGAAAAGGCACTTCTCTTTTCAGGTGGACAGAGCAGCGA 453
 QY 418 ACTGCTTGGCCCGGTACTGCTCAAAAATCCCGGAATAATTTCTGGCTGATGAACCAACC 477

Db 454 ATTGCAATTGCTCGAGCTCTGCTCAAGAAATCCAAAATTTCTTGTAGCAAGCAACA 513
QY 478 GGAGCTCTAGATTAAACAAACAGTGTAGCTAGTCATAGAAGCAATG 522
Db 514 AGTGCTCTGATGCTGAAATAGTACCTAGTGTGCAAGAAAGCTCTG 558

RESULT 12
BQ865169 759 bp mRNA linear EST 14-AUG-2002
LOCUS QGC28024.yg.abl QG ABCDI lettuce salinas Lactuca sativa cDNA clone
DEFINITION QGC28024.yg.abl QG ABCDI lettuce salinas Lactuca sativa cDNA clone
ACCESSION BQ865169
VERSION BQ865169.1 GI:22250634
KEYWORDS EST.
SOURCE Lactuca sativa
ORGANISM Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
Cichorieae; Lactuca.
1 (bases 1 to 759)
REFERENCE Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,
AUTHORS Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J.,
Ellison, P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y.,
Lai, Z., Church, S., Jackson, L. and Bradford, K.
TITLE Lettuce and Sunflower ESTs from the Compositae Genome Project
JOURNAL http://compgenomics.ucdavis.edu/
COMMENT Unpublished (2002)
Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-3659
Email: akozik@ucdavis.edu [michelmore@vegmail.ucdavis.edu]
belongs to contig QG_CA_Contig1417, see http://cgdb.ucdavis.edu/
for details.
Plate: QGC28 row: O column: 24.

FEATURES
source
1..759
Location/Qualifiers
/organism="Lactuca sativa"
/mol_type="mRNA"
/cultivar="Salinas"
/db_xref="taxon:4236"
/clone="QG28024"
/lab_host="E.coli"
/clone_lib="QG_ABCDI lettuce salinas"
/notes="Vector: pBRCNAsf1AB; The library was constructed
from 10 different sources of RNA from a single genotype.
Separate cDNAs were generated using primers that
incorporated unique 5' and 3' tags to distinguish each
source of RNA. cDNAs were then pooled, size-fractionated,
directionally cloned into a custom medium-copy vector and
transformations made with four size classes to minimize
size bias. Details of each source of RNA and library
construction can be obtained at http://cgdb.ucdavis.edu/
TAG TISSUE=chemical induction
TAG_LIB=QG ABCDI lettuce salinas
TAG_SEQ=TTGAGCCGGG"

ORIGIN
Query Match 9.9%; Score 60.4; DB 13; Length 759;
Best Local Similarity 58.2%; Pred. No. 1.3e-07;
Matches 106; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 380 AGCAGTTTTGAACTCTCTGTGGGCAACAAACAGCACTGGTTGGCCGGGTACTGC 439
Db 408 AACTGTGCTCAAGCTTAGTGGGGGTGAGAAACAAAGAGTGGCTCTGGCTCGCTTCC 467
QY 440 TCAAAATCCCGCAATTAATTTCTGGCTGATGAACCAACCGGAGCTCTAGATTTAACAAACA 499

Db 468 TGAAGCTCCTCTATTTTGTGTGATGAAGCAACAAGTGTCTCTGTATGACACACAG 527
QY 500 GTGAGCTAGTCATAGAAGCACTTGTAGAGCACTGCGCGCAAAAGGCGCCACCGTTGTTG 559
Db 528 AGGCAGAGATACTGACAGCACTTGTAGTCACTTGTGCAATAACCGCAACAGCAGTCTTTGTT 587
QY 560 CT 561
Db 588 CT 589

RESULT 13
AU090519 576 bp mRNA linear EST 02-APR-2002
LOCUS AU090519
DEFINITION AU090519 Rice mature leaf Oryza sativa (japonica cultivar-group)
CDNA clone S20717, mRNA sequence.
ACCESSION AU090519
VERSION AU090519.1 GI:7608738
KEYWORDS EST.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
1 (bases 1 to 576)
REFERENCE Sasaki, T. and Yamamoto, K.
AUTHORS Rice cDNA from mature leaf (2000)
TITLE Unpublished (2000)
JOURNAL
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsasaki@abr.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/
PROJECT "RGP"

FEATURES
source
1..576
Location/Qualifiers
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="S20717"
/tissue_type="mature leaf"
/clone_lib="Rice mature leaf"

ORIGIN
Query Match 9.8%; Score 59.4; DB 9; Length 576;
Best Local Similarity 55.6%; Pred. No. 2.5e-07;
Matches 114; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 357 TGGTCTTCAGTCGTCGAGGAAAAGCCAGTTTTTTGAACCTCTCTCGTGGCGCAACACACG 416
Db 126 TGGTATGATACTTTAGTTGGCCAGCGCGGTACTCAGCTCTCTCGAGGACAAAACACAG 185
QY 417 AACTGCGTTGGCCGGGTACTGCTCAAAAATCCCGAATTAATTCGCTGATGAACCAAC 476
Db 186 AATTGCAATTGCAAGAGCCATCCCTCAAGATCCAAAATCCCTTTTCTCGATGAAGCAAC 245
QY 477 CGGAGCTCTAGATTTAACAAACAGTGTAGTCTAGTATGAAGCACTTGAGAGCACTGCGCGA 536
Db 246 AAGTGCACCTTGATGTGGAGTCTGAGAGGATAGTTTCAGAGGCACTAAATAGATGATGCT 305
QY 537 CAAGGCGCCACCGTTGTTGCT 561
Db 306 AGAAGAACCACACTCGTTGTCGT 330

RESULT 14
H54613 358 bp mRNA linear EST 21-SEP-1995
LOCUS H54613
DEFINITION YQ92a09.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
IMAGE:203224 5' similar to SP:S21957 P-GLYCOPROTEIN PGPI -

ARABIDOPSIS ; mRNA sequence.
ACCESSION H54613
VERSION H54613.1 GI:995139
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 358)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 315
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: M13RP1
High quality sequence stop: 315.
Location/Qualifiers
1..358
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:3772355"
/db_xref="taxon:9606"
/clone="IMAGE:203224"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares fetal liver spleen INFLS"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia) with a modified polylinker; Site 1: Pac I; Site 2: Eco RI; 1st strand cDNA was primed with a Pac I...oligo(dT) primer [5' AACGGAAGAATTAATAAGATCTTTTITTTTTTTT 3'] , double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

Query Match 9.7%; Score 59.2; DB 14; Length 358;
Best Local Similarity 58.5%; Pred. No. 2.3e-07;
Matches 103; Conservative 0; Mismatches 73; Indels 0; Gaps 0;
QY 385 GTTTTGAACCTCTGTTGGCGACACACGAACTGCGTTGCGCGGCTACTGCTCAA 444
Db 126 GGTGTTCTCTCTCAGTGGCGACAAACAGCGATTCGCGTCTGCTCTAAAG 185
QY 445 AATCCCGAATAATTCGTGGTGTAGTGAACCAACCGAGCTCTAGATTAAACACAGTGAG 504
Db 186 AATCCCAAAATTCCTCTAGTAGTGAAGCAACAGTGCCTGATGCGGAATAGTAC 245
QY 505 CTAGTCATAGAGCATTCAGAGCACTGCCGACAAAGCGCCACCGTTGTTGTC 560
Db 246 CTTGTTCAAGAAGCTCTAGATCGACTGATGATGGAAGACGGTGTAGTTATGC 301

Search completed: July 30, 2004, 14:14:26
Job time : 2800 secs

VERSION BF700602.1 GI:11986010
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 426)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCW1120 row: h column: 20
High quality sequence stop: 426.
Location/Qualifiers
1..426
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4285507"
/tissue_type="Primitive neuroectoderm"
/lab_host="DH10B (TI phage-resistant)"
/clone_lib="NIH_MGC_56"
/note="Organ: Brain; Vector: pDNR-LIB (Clontech); Site 1: SfiI (ggccattatggcc); Site 2: SfiI (ggccattatggcc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CAGGCCATTATGGC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCCGAGCGCGGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

Query Match 9.7%; Score 59.2; DB 10; Length 426;
Best Local Similarity 58.5%; Pred. No. 2.5e-07;
Matches 103; Conservative 0; Mismatches 73; Indels 0; Gaps 0;
QY 385 GTTTTGAACCTCTGTTGGCGACACACGAACTGCGTTGCGCGGCTACTGCTCAA 444
Db 238 GGTGTTCTCTCTCAGTGGCGACAAACAGCGATTCGCGTCTGCTCTAAAG 297
QY 445 AATCCCGAATAATTCGTGGTGTAGTGAACCAACCGAGCTCTAGATTAAACACAGTGAG 504
Db 298 AATCCCAAAATTCCTCTAGTAGTGAAGCAACAGTGCCTGATGCGGAATAGTAC 357
QY 505 CTAGTCATAGAGCATTCAGAGCACTGCCGACAAAGCGCCACCGTTGTTGTC 560
Db 358 CTTGTTCAAGAAGCTCTAGATCGACTGATGATGGAAGACGGTGTAGTTATGC 413

Search completed: July 30, 2004, 14:14:26
Job time : 2800 secs

RESULT 15
BF700602
LOCUS BF700602 426 bp mRNA linear EST 22-DEC-2000
DEFINITION 602128776f1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4285507 5', mRNA sequence.
ACCESSION BF700602

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 27, 2004, 16:28:59 ; Search time 54 Seconds
(without alignments)
1062.169 Million cell updates/sec

Title: US-09-868-338-9
Perfect score: 1019
Sequence: 1 MEINDLKKSFGVRLWQL.....VVVATHSLFRESADTIKIL 203

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:.*
1: Geneseq1980s:.*
2: Geneseq1990s:.*
3: Geneseq2000s:.*
4: Geneseq2001s:.*
5: Geneseq2002s:.*
6: Geneseq2003as:.*
7: Geneseq2003bs:.*
8: Geneseq2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1019	100.0	203	3	AAB12592 Brevibact
2	408	40.0	213	6	ABU01071 S. pneumo
3	399	39.2	213	6	ABU02491 S. pneumo
4	399	39.2	213	6	ABP81472 Streptoco
5	386	37.9	213	6	ABU00443 S. pneumo
6	385	37.8	213	2	AAW80649 S. pneumo
7	380.5	37.3	210	6	ABU02461 S. pneumo
8	380.5	37.3	210	6	ABP81469 Streptoco
9	373	36.6	211	5	ABBS3309 Lactococc
10	362.5	35.6	465	6	ABU36156 Protein e
11	359	35.2	258	4	AAU45701 Propionib
12	359	35.2	258	6	ABM42220 Propionib
13	353	34.6	270	7	ADC97592 E. faeciu
14	353	34.6	466	6	ABU35504 Protein e
15	352	34.5	237	5	ABP26089 Streptoco
16	348	34.2	250	5	ABP26471 Streptoco
17	344.5	33.8	227	6	ABU39077 Protein e
18	344	33.8	207	5	ABB49925 Listeria
19	344	33.8	219	2	AAV34450 Porphyrom
20	344	33.8	238	2	AAV34327 Porphyrom
21	343	33.7	224	5	ABP26472 Streptoco
22	339.5	33.3	222	4	ABP76730 Corynebac
23	339.5	33.3	222	4	AAG90702 C glutami
24	339	33.3	259	5	ABBS5099 Lactococc
25	339	33.3	648	4	AAU38402 Salmonell

26	338.5	33.2	666	5	ABB47488	Listeria
27	337	33.1	252	6	ABU16155	Protein e
28	337	33.1	252	6	ABM72784	Staphyloc
29	336	33.0	236	4	AAU37889	Streptoco
30	336	33.0	249	4	AAU34146	Staphyloc
31	336	33.0	252	4	AAU37113	Staphyloc
32	333.5	32.7	209	2	AAW80648	S. pneumo
33	333.5	32.7	231	4	ABBS2924	Escherich
34	333.5	32.7	231	4	ABBS2924	Escherich
35	333	32.7	233	5	ABB48879	Listeria
36	333	32.7	236	6	ABU02123	S. pneumo
37	332	32.6	252	6	ABU43928	Protein e
38	332	32.6	256	5	ABP39755	Staphyloc
39	331.5	32.5	265	5	ABP65531	Bifidobac
40	331	32.5	234	6	ADB12657	Alloioioc
41	331	32.5	237	6	ADB12655	Alloioioc
42	331	32.5	664	5	ABBS4027	Lactococc
43	330.5	32.4	663	4	AAU36276	Pseudomon
44	330.5	32.4	663	6	ABP59941	Microbial
45	330	32.4	241	6	ABU32430	Protein e

ALIGNMENTS

RESULT 1
AAB12592 standard; protein; 203 AA.
XX AC AAB12592;
XX AC
XX 12-SEP-2003 (revised)
XX 09-NOV-2000 (first entry)
XX DE Brevibacterium lactofermentum gltD ORF2 protein SEQ ID NO:9.
XX KW Brevibacterium lactofermentum; ABC transporter; breeding; ATPase;
XX KW corynebacterium; gltBD; L-glutamic acid.
XX OS Corynebacterium glutamicum.
XX PN WO200037647-A1.
XX PD 29-JUN-2000.
XX PF 16-DEC-1999; 99WO-JP007079.
XX PR 18-DEC-1998; 98JP-00360621.
XX PA (AJIN) AJINOMOTO CO INC.
XX PI Kanno S, Kimura E, Matsui K, Nakamatsu T;
XX XX WPI; 2000-452189/39.
XX DR N-PSDB; AAA65395.
XX ABG transporter constituent of Brevibacterium lactofermentum, its encoded gene and variants, applicable in breeding Corynebacteria particularly for production of L-glutamic acid.
XX PS Claim 5; Page 30-31; 34pp; Japanese.
XX CC The present invention describes a protein (I) which can be used to construct an ATP binding cassette (ABC) transporter. ABC transporter DNAs can be used in breeding Corynebacteria particularly for production of L-glutamic acid. The present sequence represents an ORF (open reading frame) protein from the Brevibacterium lactofermentum gltBD gene, which is used in the exemplification of the present invention. (Updated on 12-SEP-2003 to standardise OS field)
XX SQ Sequence 203 AA;
Query Match 100.0%; Score 1019; DB 3; Length 203;

Best Local Similarity 100.0%; Pred. No. 1.1e-103;
Matches 203; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MIEINDLKXGVRILWQGLSHKFLPGTMTALTGASGSGKSTLLNCLGTLDPSSQILV 60
Db 1 MIEINDLKXGVRILWQGLSHKFLPGTMTALTGASGSGKSTLLNCLGTLDPSSQILV 60
Qy 61 EDVLLKLSRKQRLYRKNVTGYLFQDYALIPDRTVKFNQLAVEGKHKWPPIQVJHVG 120
Db 61 EDVLLKLSRKQRLYRKNVTGYLFQDYALIPDRTVKFNQLAVEGKHKWPPIQVJHVG 120
Qy 121 LESFEKPVFELSGEGEORTALARVILKNPRIILADEPTGALDITNSSELVIALRALADK 180
Db 121 LESFEKPVFELSGEGEORTALARVILKNPRIILADEPTGALDITNSSELVIALRALADK 180
Qy 181 GATVVVATHSPLFRESADTIKL 203
Db 181 GATVVVATHSPLFRESADTIKL 203

RESULT 2
ABU01071 ID ABU01071 standard; protein; 213 AA.
XX AC ABU01071;
XX XX
DT 23-OCT-2003 (revised)
DT 11-FEB-2003 (first entry)
XX XX
DE S. pneumoniae type 4 strain protein from coding region #642.
XX Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;
KW antinflammatory; antibacterial; immunostimulant; auditory; respiratory;
KW gene therapy; vaccine.
XX XX
OS Streptococcus pneumoniae; type 4 strain.
XX XX
FN WO200277021-A2.
XX XX
PD 03-OCT-2002.
XX XX
PF 27-MAR-2002; 2002WO-IB002163.
XX XX
PR 27-MAR-2001; 2001GB-00007658.
XX PA (CHIR-) CHIRON SPA.
XX PA (GENO-) INST GENOMIC RES.
XX PI Masignani V, Tettelin H, Fraser C;
XX XX
DR WPI; 2003-040579/03.
DR N-PSDB; ABX06354.
XX XX
PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,
PT useful as medicaments for treating or preventing a disease or infection
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
PT ear infection.
XX XX
PS Claim 1; SEQ ID NO 1284; 56pp; English.
XX XX
CC The invention relates to a protein comprising or having at least 50%
CC identity to any of the 2469 amino acid sequences, identified in the
CC specification (available on a computer readable format), or its fragment,
CC expressed from 2469 of 2489 identified DNA coding regions from the
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
CC AB556454. Also included are an antibody which binds one of the proteins,
CC treating a patient by administering the protein, DNA or antibody (in a
CC composition), a kit comprising first and second primers, which are the
CC nucleic acid cited above or fragments between nucleotides 8-100 of a
CC sequence not defined in the specification, for amplifying a target
CC sequence contained within a Streptococcus nucleic acid sequence, where
CC the first primer is substantially complementary to the target sequence
CC and the second primer is substantially complementary to the complement of

the target sequence, and where the parts of the primers having
substantial complementarity define the termini of the target sequence to
be amplified, assay comprising contacting a test compound with the
protein, and determining whether the test compound binds to the protein
and a Streptococcus pneumoniae bacterium, where one or more genes
encoding the proteins has been rendered inactive. The proteins, nucleic
acid molecules, antibody and compositions are useful as medicaments for
treating or preventing a disease or infection due to streptococcus
bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis
media or ear infection. They are also useful in developing vaccines,
diagnostics and antibiotics. The methods are useful for identifying
immunodominant proteins. The present sequence is one of the 2469 proteins
expressed by the identified coding regions from the genomic sequence.
Note: The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to
standardise OS field)
XX SQ Sequence 213 AA;
Query Match 40.0%; Score 408; DB 6; Length 213;
Best Local Similarity 41.9%; Pred. No. 3e-36;
Matches 88; Conservative 44; Mismatches 68; Indels 10; Gaps 3;
Qy 1 MIEINDLKXGVRILWQGLSHKFLPGTMTALTGASGSGKSTLLNCLGTLDPSSQILV 60
Db 1 MIEINDLKXGVRILWQGLSHKFLPGTMTALTGASGSGKSTLLNCLGTLDPSSQILV 60
Qy 61 EDVLLKLSRKQRLYRKNVTGYLFQDYALIPDRTVKFNQLAVEGKHKWPPIQV 114
Db 61 DQTSKDI---KPSVFRDYLGYLFQDFGLIESQTVKENLNLGLVGVKKKEKISLMKQ 117
Qy 115 VLHVGLESFE-EKVPFELSGEGEORTALARVILKNPRIILADEPTGALDITNSSELVIEA 173
Db 118 ALNRVNLGYLDLQPIFELSGGEAQRVALAKIILKDPFLILADEPTASLDPKNSELLSI 177
Qy 174 LRALADKATVVVATHSPLFRESADTIKL 203
Db 178 LESLKNPNRRTIIIIATHNPLIWEQVDQVIRV 207
RESULT 3
ABU02491 ID ABU02491 standard; protein; 213 AA.
XX AC ABU02491;
XX XX
DT 23-OCT-2003 (revised)
DT 11-FEB-2003 (first entry)
XX XX
DE S. pneumoniae type 4 strain protein from coding region #2069.
XX Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;
KW antinflammatory; antibacterial; immunostimulant; auditory; respiratory;
KW gene therapy; vaccine.
XX OS Streptococcus pneumoniae; type 4 strain.
XX XX
FN WO200277021-A2.
XX XX
PD 03-OCT-2002.
XX XX
PF 27-MAR-2002; 2002WO-IB002163.
XX XX
PR 27-MAR-2001; 2001GB-00007658.
XX PA (CHIR-) CHIRON SPA.
XX PA (GENO-) INST GENOMIC RES.
XX PI Masignani V, Tettelin H, Fraser C;
XX XX
DR WPI; 2003-040579/03.
DR N-PSDB; ABX07781.

XX New proteins and nucleic acid molecules from *Streptococcus pneumoniae*,
PT useful as medicaments for treating or preventing a disease or infection
PT due to *streptococcus* bacteria, such as pneumonia, sepsis, otitis media or
PT ear infection.

PS Claim 1; SEQ ID NO 4138; 56pp; English.

The invention relates to a protein comprising or having at least 50% identity to any of the 2469 amino acid sequences, identified in the specification (available on a computer readable format), or its fragment, expressed from 2469 of 2489 identified DNA coding regions from the *Streptococcus pneumoniae* type 4 strain genomic sequence appearing as AS55454. Also included are an antibody which binds one of the proteins, treating a patient by administering the protein, DNA or antibody (in a composition), a kit comprising first and second primers, which are the nucleic acid cited above or fragments between nucleotides 8-100 of a sequence not defined in the specification, for amplifying a target sequence contained within a *Streptococcus* nucleic acid sequence, where the first primer is substantially complementary to the target sequence and the second primer is substantially complementary to the complement of the target sequence, and where the parts of the primers having substantial complementarity define the termini of the target sequence to be amplified, assay comprising contacting a test compound with the protein, and determining whether the test compound binds to the protein and a *Streptococcus pneumoniae* bacterium, where one or more genes encoding the proteins has been rendered inactive. The proteins, nucleic acid molecules, antibody and compositions are useful as medicaments for treating or preventing a disease or infection due to *Streptococcus* bacteria, particularly *S. pneumoniae*, such as pneumonia, sepsis, otitis media or ear infection. They are also useful in developing vaccines, immunomodulators and antibiotics. The methods are useful for identifying immunodominant proteins. The present sequence is one of the 2469 proteins expressed by the identified coding regions from the genomic sequence. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences). (Updated on 23-OCT-2003 to standardise OS field)

KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;
 KW gene therapy; vaccine.
 XX
 OS Streptococcus pneumoniae; type 4 strain.
 XX
 FN WO200277021-A2.
 XX
 XX
 PD 03-OCT-2002.
 XX
 FF 27-MAR-2002; 2002WO-IB002163.
 XX
 PR 27-MAR-2001; 2001GB-00007658.
 XX
 PA (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Masigani V, Tettelin H, Fraser C;
 XX
 DR WPI; 2003-040579/03.
 DR N-PSDB; ABX05722.
 XX
 PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,
 PT useful as medicaments for treating or preventing a disease or infection
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
 PT ear infection.
 XX
 PS Claim 1; SEQ ID NO 20; 56pp; English.
 XX
 CC The invention relates to a protein comprising or having at least 50%
 CC identity to any of the 249 amino acid sequences, identified in the
 CC specification (available on a computer readable format), or its fragment,
 CC expressed from 2469 of 2489 identified DNA coding regions from the
 CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
 CC AB356454. Also included are an antibody which binds one of the proteins,
 CC treating a patient by administering the protein, DNA or antibody (in a
 CC composition), a kit comprising first and second primers, which are the
 CC nucleic acid cited above or fragments between nucleotides 8-100 of a
 CC sequence not defined in the specification, for amplifying a target
 CC sequence contained within a Streptococcus nucleic acid sequence, where
 CC the first primer is substantially complementary to the target sequence
 CC and the second primer is substantially complementary to the complement of
 CC the target sequence, and where the parts of the primers having
 CC substantial complementarity define the termini of the target sequence to
 CC be amplified, assay comprising contacting a test compound with the
 CC protein, and determining whether the test compound binds to the protein
 CC encoding the proteins has been rendered inactive. The proteins, nucleic
 CC acid molecules, antibody and compositions are useful as medicaments for
 CC treating or preventing a disease or infection due to streptococcus
 CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis
 CC media or ear infection. They are also useful in developing vaccines,
 CC diagnostics and antibiotics. The methods are useful for identifying
 CC immunodominant proteins. The present sequence is one of the 2469 proteins
 CC expressed by the identified coding regions from the genomic sequence.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to
 CC standardise OS field)
 XX
 SQ Sequence 213 AA;
 Query Match 37.9%; Score 386; DB 6; Length 213;
 Best Local Similarity 40.2%; Pred. No. 7.9e-34;
 Matches 84; Conservative 45; Mismatches 74; Indels 6; Gaps 1;
 QY 1 MIEINDLKSGFVRLWQGLSHKFLPGTWTALTGASGSGKSTLNLGLDKPSSGQILV 60
 DB 1 MIELKNITKTGGKVLIDNLNLSRIDQDGLVAIVKSGSGKSTLNLGLDGDYSGRYEI 60
 QY 61 EDVLLKLSRQRLYRKNTVGYLFQDYALIPDRYKFNQLAV-----EKHKWPEIPQ 114
 DB 61 FGQTNLAVNSAKSQTIREHISYLFQNFALIDBETVEYNLMALKYVKLPKDKLKKVEE 120

QY 115 VHAIVGLSEFEKPFVFLSGGCGQORTALRVLLKNPRIILADEPTGALDNLNSLVIEAL 174
 DB 121 ILERVGLSATLHORVSELSGEGQRIAVAILKPSQLILADEPTGSLDPENRDLVKEL 180
 QY 175 RALADKGTATVVVATHSPFLRESADTIKL 203
 DB 181 LEMNREGXTVIIIVTHDAYVAQCCHRIEL 209
 RESULT 6
 AAW80649
 ID AAW80649 standard; protein; 213 AA.
 XX
 AC AAW80649;
 XX
 DT 24-DEC-1998 (first entry)
 XX
 DE S. pneumoniae transport protein.
 XX
 KW Streptococcus pneumoniae protein; recombinant; gene expression; DNA chip;
 KW virulence; antibody; infection; detection; treatment; hypothetical;
 KW cell wall biosynthetic, external target; minimal gene set protein.
 XX
 OS Streptococcus pneumoniae.
 XX
 FN WO9826072-A1.
 XX
 PD 18-JUN-1998.
 XX
 PF 09-DEC-1997; 97WO-US022578.
 XX
 PR 13-DEC-1996; 96US-0036281P.
 XX
 PA (ELIL) LILLY & CO ELI.
 XX
 PI Baltz RH, Burgett SG, Dehoff BS, Hoskins JA, Jaskunas SR;
 PI Mills BJ, Norris FH, Peery RB, Rostek PK, Rostek PR, Skatrud PL;
 PI Smith MC, Solenberg PJ, Treadway PJ, Young Bellido ML;
 XX
 DR WPI; 1998-348529/30.
 DR N-PSDB; AAW65223.
 XX
 PT Streptococcus pneumoniae nucleic acid sequences - used in DNA chips for
 PT evaluating gene expression, and identification of virulence genes.
 XX
 PS Claim 3; Page 236-237; 333pp; English.
 XX
 CC This sequence represents a Streptococcus pneumoniae transport protein.
 CC The invention provides DNA sequences (AAV65201 to AAV65304) from the
 CC Streptococcus pneumoniae genome and corresponding protein sequences
 CC (AAW6505 to AAW6528). The protein sequences are classified as
 CC hypothetical, cell wall biosynthetic, external target, or minimal gene
 CC set proteins. A recombinant host containing a vector comprising any of
 CC the above nucleic acids can be used for the recombinant expression of the
 CC proteins. The invention also provides a DNA chip having arrayed on it at
 CC least 15 base pair fragment of any one or more of these DNA sequences.
 CC The DNA chip can be used methods for evaluating gene expression in S.
 CC pneumoniae and for identifying virulence genes in S. pneumoniae.
 CC Antibodies that selectively bind to the above proteins or peptide
 CC fragments can be used to treat S. pneumoniae infection. The antibodies
 CC can also be used to detect S. pneumoniae cells
 XX
 SQ Sequence 213 AA;
 Query Match 37.8%; Score 385; DB 2; Length 213;
 Best Local Similarity 39.7%; Pred. No. 1e-33;
 Matches 83; Conservative 46; Mismatches 74; Indels 6; Gaps 1;
 QY 1 MIEINDLKSGFVRLWQGLSHKFLPGTWTALTGASGSGKSTLNLGLDKPSSGQILV 60
 DB 1 MIELKNITKTGGKVLIDNLNLSRIDQDGLVAIVKSGSGKSTLNLGLDGDYSGRYEI 60
 QY 61 EDVLLKLSRQRLYRKNTVGYLFQDYALIPDRYKFNQLAV-----EKHKWPEIPQ 114

Db 61 FQQTNLAVNSAKSGTIIREHISYLFQNFALIDDETVEYNLMALKVVKPKDKLKKVZE 120
 QY 115 VLHVGLESFEKVPFELSGEQORTALARVLKNPRIILADPTGALDITNSLVEIAL 174
 Db 121 ILERVGLSATLHORVSELSGEQORAVARAILKPSQLILADEPTGSLDPENRDLVKEL 180
 QY 175 RALADKCATVVVATHSPFPRESADTIKL 203
 Db 181 LEWNRGKTVIIVTHDAYVAQQCHRVIEL 209

RESULT 7
 ABU02461
 ID ABU02461 standard; protein; 210 AA.

AC ABU02461;
 XX
 DT 23-OCT-2003 (revised)
 DT 11-FEB-2003 (first entry)
 XX
 XX S. pneumoniae type 4 strain protein from coding region #2039.
 XX Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;
 KW antinflammatory; antibacterial; immunostimulant; auditory; respiratory;
 KW gene therapy; vaccine.
 XX
 XX Streptococcus pneumoniae; type 4 strain.
 XX
 XX WO200277021-A2.
 XX
 XX 03-OCT-2002..

XX 27-MAR-2002; 2002WO-IB002163.

XX 27-MAR-2001; 2001GB-00007658.

XX (CHIR-) CHIRON SPA.

XX (GENO-) INST GENOMIC RES.

XX Masignani V, Tettelin H, Fraser C;

XX WPI; 2003-040579/03.

XX N-PSDB; ABX07751.

XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,
 PT useful as medicaments for treating or preventing a disease or infection
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
 PT ear infection.

XX Claim 1; SEQ ID NO 4078; 56pp; English.

XX The invention relates to a protein comprising or having at least 50%
 CC identity to any of the 2469 amino acid sequences, identified in the
 CC specification (available on a computer readable format), or its fragment,
 CC expressed from 2469 of 2489 identified DNA coding regions from the
 CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
 CC AS556454. Also included are an antibody which binds one of the proteins,
 CC treating a patient by administering the protein, DNA or antibody (in a
 CC composition), a kit comprising first and second primers, which are the
 CC nucleic acid cited above or fragments between nucleotides 8-100 of a
 CC sequence not defined in the specification, for amplifying a target
 CC sequence contained within a Streptococcus nucleic acid sequence where
 CC the first primer is substantially complementary to the target sequence
 CC and the second primer is substantially complementary to the complement of
 CC the target sequence, and where the parts of the primers having
 CC substantial complementarity define the termini of the target sequence to
 CC be amplified, assay comprising contacting a test compound with the
 CC protein, and determining whether the test compound binds to the protein
 CC and a Streptococcus pneumoniae bacterium, where one or more genes
 CC encoding the proteins has been rendered inactive. The proteins, nucleic
 CC acid molecules, antibody and compositions are useful as medicaments for
 CC treating or preventing a disease or infection due to streptococcus

CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis
 CC media or ear infection. They are also useful in developing vaccines,
 CC diagnostics and antibiotics. The methods are useful for identifying
 CC immunodominant proteins. The present sequence is one of the 2469 proteins
 CC expressed by the identified coding regions from the genomic sequence.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to
 CC standardise OS field)

XX Sequence 210 AA;

Query Match 37.3%; Score 380.5; DB 6; Length 210;
 Best Local Similarity 41.0%; Pred. No. 3.le-33;
 Matches 86; Conservative 41; Mismatches 72; Indels 11; Gaps 4;

QY 1 MIEINDLKSPGVRIIMWGLSHKFLPGMTALTGASGSGKSTLLNCLGTLKPSGGQLV 60
 Db 1 MIELKQVSKSGFERELFNSLMTPEAGKVYALIGSGSGKSTLLNMIGKLE-PYDGTIFY 59
 QY 61 EDVDLLKLSRQRLYRKNTVGYLFQDYALIPDRTVKFNQLAVEKHKWPEI-----PQ 114
 Db 60 RGKD---LANYKSSDFFRHELGYLEFQNGLIENQSIENKLGILIGKLSRSEORLRKQ 116
 QY 115 VLHVGLESFE-EKVPFELSGEQORTALARVLKNPRIILADPTGALDITNSLVEIA 173
 Db 117 ALEQVGLVYLDLKRIFELSGESQORVALAKIILKNPPFILADEPTASIDPATSQLIMEI 176
 QY 174 LEADKCATVVVATHSPFPRESADTIKL 203
 Db 177 LLSURDDNRLLIIITHNPFIWEMADEVPTM 206

RESULT 8

ABP81469

ID ABP81469 standard; protein; 210 AA.

XX

AC ABP81469;

XX

DT 04-MAR-2003 (first entry)

XX

XX Streptococcus pneumoniae polypeptide SEQ ID NO 386.

XX

XX Streptococcus pneumoniae; infection; otitis media; antibacterial;
 KW diagnosis; gene therapy.

XX

XX Streptococcus pneumoniae.

XX

XX WO200283855-A2.

XX

XX 24-OCT-2002.

XX

XX 12-APR-2002; 2002WO-US011524.

XX

XX 16-APR-2001; 2001US-0283948P.

XX

XX 18-APR-2001; 2001US-0284443P.

XX

XX (AMCY) AMERICAN CYANAMID CO.

XX

XX Zagursky RJ, Masi AW, Green BA, Chakravarti DN, Russell DP;

XX

XX Wooters JL;

XX

XX WPI; 2003-093010/08.

XX

XX N-PSDB; ABZ42317.

XX

XX New Streptococcus pneumoniae polynucleotides, useful for treating or
 PT preventing S. pneumoniae infections, or non-systemic diseases, e.g.
 PT otitis media, which are induced or exacerbated by S. pneumoniae.

XX

XX Claim 42; Page 637-638; 1091pp; English.

XX

XX The invention relates to isolated polynucleotides (ABZ72147-ABZ42522) of
 CC a Streptococcus pneumoniae genomic sequence, a fragment or degenerate

CC variant of the polynucleotide or a nucleic acid sequence 95% identical to
 CC one of the polynucleotides. The S. pneumoniae polynucleotides and encoded
 CC polypeptides (ABP81299-ABP81674) are useful for treating or preventing S.
 CC pneumoniae infections or non-systemic diseases, e.g. otitis media, which
 CC are induced or exacerbated by S. pneumoniae. These are also useful for
 CC detecting S. pneumoniae in a biological sample or diagnosing S.
 CC pneumoniae infection in a subject. The polynucleotides have antibacterial
 CC activity and are useful in gene therapy

XX Sequence 210 AA;

Query Match 37.3%; Score 380.5; DB 6; Length 210;
 Best Local Similarity 41.0%; Pred. No. 3.1e-33;
 Matches 86; Conservative 41; Mismatches 72; Indels 11; Gaps 4;
 QY 1 MIEINDLKSGFVRIWQGLSHKFLPGTMTALTGASGSGKSTLNLCLGTLDKPSGGILV 60
 DB 1 MIEIKQVSKFGRELSNLSMTPEAGKVVALIGSSGSGKTLNMIGKLE-PYDTIFY 59
 QY 61 EDVDLLKSLTRKQRLYKNTVGYLFQDYALIPDRTVKFNQLQAVEKHKWPFI-----PQ 114
 DB 60 RGKD---LANYKSDFFRHELGVLFGFNGLIENQSIENLKLGLIGCKLSRSRQRLRQK 116
 QY 115 VLHVGLESPE-EKVPFELSGGEQRTALARVLLKNPRIILADEPTGALDLTNSLVIEA 173
 DB 117 ALEQVGLVLDLRIRFELSGGESQRTALAKIKLNPPFIILADEPTASIDPATISQLIMEI 176
 QY 174 LRALADKGVTVVATHSPFRESADTIKL 203
 DB 177 LLSLRDNDRLIIITHNPATWMADEVFTM 206

RESULT 9

ABB53309
 ID ABB53309 standard; protein; 211 AA.

XX ABB53309;

XX 29-AUG-2003 (revised)

DT 16-MAY-2002 (first entry)

XX Lactococcus lactis protein yabE.

XX Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.

XX Lactococcus lactis; IL1403.

XX FR2807446-A1.

XX 12-OCT-2001.

XX 11-APR-2000; 2000FR-00004630.

XX 11-APR-2000; 2000FR-00004630.

XX (INRG) INRA INST NAT RECH AGRONOMIQUE.

XX Bolotine A, Sorokine A, Renault P, Ehrlich SD;

XX WPI; 2002-043418/06.

XX New nucleotide sequence useful in the identification of Lactococcus
 PT lactis and related species.

PS Claim 6; SEQ ID NO 11; 2504pp; French.

XX The present invention is related to a Lactococcus lactis nucleotide
 CC sequence (ABA90521) and related proteins (ABB53300-ABB55621). The nucleic
 CC acid sequence is useful in the detection and/or amplification of nucleic
 CC acid sequence, particularly to identify Lactococcus lactis or related
 CC species. The proteins of the invention are useful for the biosynthesis or
 CC biodegradation of a composition of interest. The invention helps research
 CC in lactic bacteria, particularly useful in the production of yogurt and

CC cheese. Note: The sequence data for this patent is based on equivalent
 CC patent WO200177334 (published 18-OCT-2001) which is available in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences. (Updated on 29-AUG-2003 to
 CC standardise OS field)

XX Sequence 211 AA;

Query Match 36.6%; Score 373; DB 5; Length 211;
 Best Local Similarity 39.2%; Pred. No. 2.1e-32;
 Matches 82; Conservative 42; Mismatches 79; Indels 6; Gaps 1;
 QY 1 MIEINDLKSGVRIWQGLSHKFLPGTMTALTGASGSGKSTLNLCLGTLDKPSGGILV 60
 DB 1 MIEIBELTKSYKGIHIFDKLNLRIPEGKQTAIYGTSGAGKSTLNLIGLIEDYDGGKYF 60
 QY 61 EDVDLLKSLTRKQRLYKNTVGYLFQDYALIPDRTVKFNQLQAV-----EKHKWPFIQ 114
 DB 61 NQGFAPPFNSSIALKQRENKISYLFQNFALLEDETIEKNLSIALIYGRISKKEKKMKK 120
 QY 115 VLHVGLESPEKVPFELSGGEQRTALARVLLKNPRIILADEPTGALDLTNSLVIEAL 174
 DB 121 LLLQVGNHRLNTKVYSLSGGKQRTAARALLKESQLILADEPTGSLDTENRNEVIAL 180
 QY 175 RALADKGVTVVATHSPFRESADTIKL 203
 DB 181 RQEVDKGKAVIVTHDSYLKVSVDLVEI 209

RESULT 10

ABU36156

ID ABU36156 standard; protein; 465 AA.

XX AC ABU36156;

XX 19-JUN-2003 (first entry)

DE Protein encoded by Prokaryotic essential gene #21683.

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX Mycoplasma pneumoniae.

XX WO200277183-A2.

XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

DR N-PSDB; ACA40026.

XX New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 64080; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid

CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 465 AA;

Query Match 35.6%; Score 362.5; DB 6; Length 465;
 Best Local Similarity 36.7%; Pred. No. 9.9e-31;
 Matches 77; Conservative 54; Mismatches 70; Indels 9; Gaps 4;
 QY 1 MIBINDLKK--SGV--RILWQGLSHKFLPGTWTALTGASGSKSTLLNCLGTLKPPSG 56
 Db 231 IIEELKNVYKTYINGVTNVAIKGIDILKAHDPIVILGPGSGKTLNIIISGMDRPPSG 290
 QY 57 QILVEDVDLLKSTRKQRYKNTVGYLFQDYALIPDRYKFNQLQAVEKHKEWPB---I 112
 Db 291 SVVNVQEMICMDRQLTFRNRYVYIFQYGLLENLVRENVGVANLQNPDKRNI 350
 QY 113 PQVLHVGLESFEEKPVFELSGEQQORTALARVLLKNPRIILADEPTGALDNLNSELVIE 172
 Db 351 DELEAVGMKHLQKLPNELSGQQQQRVSTARAFAXNPILLIFGDEPTGALDLEMTQIVLK 410
 QY 173 ALPALADK-GATVVVATHSPFPRESADTII 201
 Db 411 QFLAKQRYKTTWIVTHNNLIAQLADLVI 440

RESULT 11
 AAU45701
 ID AAU45701 standard; protein; 258 AA.
 XX
 AC AAU45701;
 XX
 DT 27-FEB-2002 (first entry)
 XX
 DE Propionibacterium acnes immunogenic protein #6597.
 XX
 KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KW dermatological; osteopathic; neuroprotectant.
 XX
 OS Propionibacterium acnes.
 XX
 FN WO200181581-A2.
 XX
 PD 01-NOV-2001.
 XX
 PF 20-APR-2001; 2001WO-US012865.

XX 21-APR-2000; 2000US-0199047P.
 PR 02-JUN-2000; 2000US-0208841P.
 BR 07-JUL-2000; 2000US-0216747P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
 DR WPI: 2001-616774/71.
 DR N-PSDB; AAS59526.
 XX
 PT Propionibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris.
 XX
 PS Claim 3; SEQ ID NO 6896; 1069pp; English.
 XX
 CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
 CC polypeptides. The proteins and their associated DNA sequences are used in
 CC the treatment, prevention and diagnosis of medical conditions caused by
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
 CC P. acnes is also involved in infections of bone, joints and the central
 CC nervous system, however it is particularly involved in the inflammatory
 CC lesions associated with acne vulgaris. A method for detecting the
 CC presence or absence of P. acnes in a patient comprises contacting a
 CC sample with a binding agent that binds to the proteins of the invention
 CC and determining the amount of bound protein in the sample. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC specific for P. acnes proteins. These antibodies can be used to
 CC downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
 CC this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 258 AA;

Query Match 35.2%; Score 359; DB 4; Length 258;
 Best Local Similarity 40.5%; Pred. No. 9.9e-31;
 Matches 85; Conservative 36; Mismatches 65; Indels 24; Gaps 3;
 QY 12 GVRILWQGLSHKP-----LPGTWTALTGASGSKSTLLNCLGTLKPPS 54
 Db 34 GGSIVAGSVSHNFSTTHNRNLVLDVTVSMQPGVMTLLHGVSGSGKTLNIIISGMLKPT 93
 QY 55 SGQILVEDVDLLKSTRKQRYKNTVGYLFQDYALIPDRYKFNQLQAVEKHKEWPBIPQ 114
 Db 94 SGTVFEGSDIYSLSTSERODIRLNRIQMIQFQHSLLIVDFTVRENVELLARVFGSGRSR 153
 QY 115 V-----LHVGLESFEEKPVFELSGEQQORTALARVLLKNPRIILADEPTGALDNLNSEL 169
 Db 154 VMAIEALERVGIAHLQDPRQLSGGQAQVGIARAIAGDRPILLADEPTGALDRRSQM 213
 QY 170 VIEALRALA--DKGATVVVATHSPFPRESA 197
 Db 214 VFELLRALADEKQRTVLLSSHDPTAKYA 243

RESULT 12
 ABM42220
 ID ABM42220 standard; protein; 258 AA.
 XX
 AC ABM42220;
 XX
 DT 20-OCT-2003 (first entry)
 XX
 DE Propionibacterium acnes transporter-related polypeptide #6896.

KW Acne vulgaris; antiseborrheic; dermatological; antibacterial;
KW immunostimulant; immune response; vaccine.

OS Propionibacterium acnes.

XX WO2003033515-A1.

XX 24-APR-2003.

XX 11-OCT-2002; 2002WO-US032727.

XX 15-OCT-2001; 2001US-00978825.

XX (CORI-) CORIXA CORP.

XX Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;

PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;

PI Barth B, Vallie-Douglas J;

XX WPI: 2003-381789/36.

DR N-PSDB; ACF64455.

XX New Propionibacterium acnes polypeptides and polynucleotides encoding the
PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
PT or for stimulating an immune response specific for a P. acnes protein.

XX Claim 3k; SEQ ID NO 6896; 1481pp; English.

XX The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
CC encoding a Propionibacterium acnes protein. The invention also relates to
CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
CC immunogenic fragments of P. acnes polypeptides. The invention
CC additionally encompasses expression vectors and host cells comprising a
CC polynucleotide of the invention; antibodies against polypeptides of the
CC invention; fusion proteins comprising a polypeptide of the invention; a
CC method for stimulating an immune response specific for a P. acnes
CC polypeptide and an isolated T cell population comprising T cells prepared
CC via this method; a vaccine composition (comprising P. acnes polypeptides,
CC polynucleotides, antibodies, fusion proteins, T cell populations, or
CC antigen-presenting cells that express the polypeptide); a method and kit
CC for detecting or determining the presence or absence of P. acnes in a
CC patient; and a method for inhibiting the development of P. acnes in a
CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
CC proteins, T cell populations or antigen-presenting cells that express the
CC polypeptides are useful for diagnosing, preventing or treating acne
CC vulgaris, or for stimulating an immune response specific for a P. acnes
CC protein. The polynucleotides can also be used as probes or primers for
CC nucleic acid hybridisation. The vaccine composition is useful for the
CC stimulation of an immune response against P. acnes, or for treating acne,
CC and the kit is useful for performing a diagnostic assay. The present
CC sequence represents a specifically claimed polypeptide which is predicted
CC to be encoded by an ORF (open reading frame) contained within the P.
CC acnes polynucleotides of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pcr_sequences

XX Sequence 258 AA;

Query Match 35.2%; Score 359; DB 6; Length 258;

Best Local Similarity 40.5%; Pred. No. 9.9e-31;

Matches 85; Conservative 36; Mismatches 65; Indels 24; Gaps 3;

QY 12 GVRILWQLSHKF-----LPGTWLTALTGASGSKTLNCLGTLDKPS 54

DB 34 GGSIVASGVSHNFSTHNRNLVLDDVTVMQPGVWTLHVGSGKTLINMGSLKPT 93

QY 55 SQQLIVEDVLLKLSRQRLRYKNTVGYLPQDYALIPDRTPKFNQLAVEKHKWPEIQ 114

DB 94 SGTVFEGSDIYLSSTSRDRLNRIGMIFQHSRLIVDFTVRENVELILRVGFGRSR 153

QY 115 V-----LHVGLEFEEKPFVLSGGECQRTALARVLLKPRILLADEPTGADLTNSEL 169

DB 154 VMATEALERVGIAHLQDRYPRQLSGGEARVGIARAIAGDRPILLADEPTGQLDRRNSQ 213

QY 170 VTEALRALA--DKGATVVVATHSPLFRESA 197

DB 214 VFELRLAALAEDEKRGRTVVLSHSDPTAKAYA 243

RESULT 13

ADC97592

ID ADC97592 standard; protein; 270 AA.

XX AC ADC97592;

XX DT 01-JAN-2004 (first entry)

XX E. faecium protein sequence SEQ ID 7219.

XX Vaccine; urinary tract infection; bacteraemia; endocarditis; wound;

KW abdominal-pelvic infection.

XX Enterococcus faecium.

XX US6583275-B1.

XX PD 24-JUN-2003.

XX PF 30-JUN-1998; 98US-00107532.

XX PR 02-JUL-1997; 97US-0051571P.

XX PR 14-MAY-1998; 98US-0085598P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Doucette-Stamm LA, Bush D;

XX WPI: 2003-799836/75.

DR N-PSDB; ADC93938.

XX New isolated nucleic acid derived from Enterococcus faecium encoding an
PT Enterococcus faecium polypeptide useful for detection, prevention and
PT treatment of a pathological condition resulting from a bacterial
PT infection.

XX Example 1; SEQ ID NO 7219; 243pp; English.

XX The invention relates to an isolated nucleic acid derived from
CC Enterococcus faecium encoding an Enterococcus faecium polypeptide having
CC one of 10 fully defined sequences given in the (or comprising 40
CC sequential nucleotides chosen from any of the nucleic acids, its
CC complement or sequences hybridising to it). Also included are a
CC recombinant vector comprising the nucleic acid operably linked to a
CC transcription regulatory element, a cell comprising the vector and a
CC single-stranded probe comprising the nucleic acid. The nucleic acids are
CC chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.
CC The nucleic acids are useful for diagnosing pathological conditions
CC resulting from E. faecium bacterial infection (e.g. urinary tract
CC infection, bacteraemia, endocarditis, wounds and abdominal-pelvic
CC infection) and for screening drugs such as agonists and antagonists. The
CC nucleic acid is useful for recombinant production of Candida albicans -
CC derived peptides or antisense polypeptides. Pharmaceutical compositions
CC and vaccines containing the nucleic acid are useful for preventing or
CC treating Enterococcus faecium infections. The present sequence represents
CC one if the disclosed E. faecium proteins.

XX Sequence 270 AA;

Query Match 34.6%; Score 353; DB 7; Length 270;

Best Local Similarity 40.9%; Pred. No. 4.9e-30;

Matches 76; Conservative 36; Mismatches 72; Indels 2; Gaps 2;

QY 20 LSHKFLPTMTALTGASGSKSTLNLCLGTLDKPSSQQLIVEDVLLKLSRQRLRYK 79

DB 63 ISFSIEKSELVILGFGAGKSTILNLGGMIDTDEGQIIDDITDIAQFSKQLTAYERT 122

QY 80 TVGVLFQDYALIPDRITVKNLQAVKHKWPEIP-OVLHAVGLSEFEKVPVLSGGEQ 138
Db 123 DVGFEVQFYNLVNLTAKENVELATEVSPDALDPVEVLQVGLAHLNFPQLSGGEQ 182
QY 139 RTALARVLLKNPRIILADEPTGALDITNSELVIEALR-ALADKATVAVVATHSPVLFRESA 197
Db 183 RVSGIRALAKNPKLLLCDEPTGALDPETGQVCLKLQNASRQHGNTVLIITHSALAPIA 242
QY 198 DTIIKL 203
Db 243 DRVHI 248

RESULT 14
ABU35504
ID ABU35504 standard; protein; 466 AA.
XX AC ABU35504;
XX 19-JUN-2003 (first entry)
XX DE Protein encoded by Prokaryotic essential gene #21031.
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX OS Mycoplasma genitalium.
XX PN WO200277183-A2.
XX PD 03-OCT-2002.
XX PF 21-MAR-2002; 2002WO-US009107.
XX PR 21-MAR-2001; 2001US-00815242.
XX PR 06-SEP-2001; 2001US-00948993.
XX PR 25-OCT-2001; 2001US-0342323P.
XX PR 08-FEB-2002; 2002US-00072851.
XX PR 06-MAR-2002; 2002US-0362699P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
XX DR N-PSDB; ACA39374.
XX PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX FS Claim 25; SEQ ID NO 63428; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent

CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences

XX Sequence 466 AA;
Query Match 34.6%; Score 353; DB 6; Length 466;
Best Local Similarity 39.2%; Pred. No. 1.1e-29;
Matches 73; Conservative 42; Mismatches 61; Indels 10; Gaps 3;
QY 21 SHKELPGTMTALTGASGSGKTLNCLGTLDKPSGGQILVEDVLLKLSTKQSLYKNT 80
Db 261 SHDFI-----VILGPGSGKTTLLNIISGMDRASSGSVIVNGYNICLNDKRLTKPQKY 315
QY 81 VGYLFQDYALIPDRITVKNLQAVKHKWP-----EIPQVLHAVGLSEFEKVPVLSGGE 136
Db 316 VGYTFQYGLPNLTVRENIIGANLQDPFSKRISIDALLEAVGMDSLQKLPNELSGGQ 375
QY 137 QORTALARVLLKNPRIILADEPTGALDITNSELVIEALRALADK-GATVVVATHSPVLFRE 195
Db 376 QQRVSIARAFANKPELLIFGDEPTGALDLEMTQIVLKQFLAKKXYQTTMIIVTHNNLIAN 435
QY 196 SADTII 201
Db 436 LADLVI 441

RESULT 15
ABP26089
ID ABP26089 standard; protein; 237 AA.
XX AC ABP26089;
XX DT 02-JUL-2002 (first entry)
XX DE Streptococcus polypeptide SEQ ID NO 1354.
XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX group A streptococcus; Streptococcus pyogenes; antibacterial;
XX antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX OS Streptococcus pyogenes.
XX PN WO200234771-A2.
XX PD 02-MAY-2002.
XX PF 29-OCT-2001; 2001WO-GB004789.
XX PR 27-OCT-2000; 2000GB-00026333.
XX PR 24-NOV-2000; 2000GB-00028727.
XX PR 07-MAR-2001; 2001GB-00005640.
XX PA (CHIR-) CHIRON SPA.
XX PA (GENO-) INST GENOMIC RES.
XX PI Telford J, Massignani V, Margarit Y Rosi, Grandi G, Fraser C;
PI Tetelin H;
XX DR WPI; 2002-352536/38.
XX DR N-PSDB; ABN66720.
XX PT New Streptococcus protein for the treatment or prevention of infection or
PT disease caused by Streptococcus bacteria, such as meningitis, and for

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 27, 2004, 16:34:05 ; Search time 19 Seconds
(without alignments)
551.583 Million cell updates/sec

Title: US-09-868-338-9
Perfect score: 1019
Sequence: 1 MIBINDLKKXSGVRLWQGL.....VVVATHSPFLRESADTIKL 203

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	353	34.6	270	4	US-09-107-532A-7219
2	332	32.6	256	4	US-09-134-001C-4600
3	330.5	32.4	1049	4	US-09-252-991A-25336
4	329	32.3	672	4	US-09-489-039A-12101
5	324	31.8	257	4	US-09-134-000C-3925
6	322.5	31.6	246	4	US-09-627-376-15
7	321	31.5	653	4	US-09-543-681A-5411
8	320.5	31.5	239	4	US-09-134-000C-6167
9	315	30.9	226	4	US-09-134-000C-5979
10	308	30.2	286	4	US-09-540-236-2931
11	306.5	30.1	276	4	US-09-489-039A-13021
12	306	30.0	242	4	US-09-134-001C-3932
13	305.5	30.0	221	4	US-09-107-532A-6667
14	305.5	30.0	221	4	US-09-134-000C-6517
15	303	29.7	396	4	US-09-489-039A-10491
16	301.5	29.6	236	4	US-09-134-001C-4595
17	300.5	29.5	287	4	US-09-489-039A-7596
18	299.5	29.4	270	4	US-09-134-001C-5561
19	299	29.3	272	4	US-09-489-039A-9558
20	298	29.2	794	4	US-09-134-000C-5518
21	295	28.9	254	4	US-09-134-001C-4440
22	295	28.9	674	4	US-09-328-352-8094
23	294.5	28.9	262	4	US-09-134-000C-5469
24	294.5	28.9	274	4	US-09-252-991A-23361
25	294	28.9	255	4	US-09-107-532A-5335
26	292	28.7	244	4	US-08-919-573-2
27	292	28.7	244	4	US-08-919-573-4

28	292	28.7	350	4	US-09-107-532A-6978
29	292	28.7	360	4	US-09-543-681A-6466
30	291	28.6	231	4	US-09-489-039A-10973
31	288.5	28.3	215	4	US-09-305-984-64
32	288.5	28.3	215	4	US-09-493-940-64
33	288.5	28.3	245	4	US-09-252-991A-20657
34	288	28.3	231	4	US-09-543-681A-6078
35	288	28.3	248	4	US-09-134-001C-3731
36	287.5	28.2	257	4	US-09-543-681A-7794
37	287.5	28.2	266	4	US-09-107-532A-5226
38	284.5	27.9	215	4	US-09-305-984-18
39	284.5	27.9	215	4	US-09-073-541A-18
40	284.5	27.9	215	4	US-09-493-940-18
41	284	27.9	247	4	US-09-107-532A-4327
42	283.5	27.8	224	4	US-09-107-532A-4444
43	283	27.8	252	4	US-09-489-039A-14256
44	282	27.7	224	4	US-09-305-984-72
45	282	27.7	224	4	US-09-305-984-74

ALIGNMENTS

RESULT 1

US-09-107-532A-7219
; Sequence 7219, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESS: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354

COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 7219:
SEQUENCE CHARACTERISTICS:
LENGTH: 270 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...270
SEQUENCE DESCRIPTION: SEQ ID NO: 7219:

US-09-107-532A-7219

[illegible]

Db 209 VMAILHQLKAOGHVTIIVTHDPQVAAQAERIVEI 242

RESULT 5

US-09-134-000C-3925

; Sequence 3925, Application US/09134000C

; Patent No. 6617156

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 032796-032

; CURRENT APPLICATION NUMBER: US/09/134,000C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/055,778

; PRIOR FILING DATE: 1997-08-15

; NUMBER OF SEQ ID NOS: 6812

; SOFTWARE: Patent in version 3.1

; SEQ ID NO 3925

; LENGTH: 257

; TYPE: PRT

; ORGANISM: Enterococcus faecalis

US-09-134-000C-3925

Query Match 31.8%; Score 324; DB 4; Length 257;

Best Local Similarity 35.1%; Pred. No. 5.3e-29;

Matches 72; Conservative 52; Mismatches 65; Indels 16; Gaps 4;

QY 1 MIEINDLKKSFGVRLWQ---GLSHKFLPGTWTALTGASGSKSTLLNCLGTLDPKSSG 57

Db 9 MLEVQNLKKVYGNIEIKYEALKGINTVQDGEFIMGPSGSKSTLLNLLATIDPTDGE 68

QY 58 ILVEDVDLLKLSRKORLYRKNVTGVLFQDYALIPDRTVKFNL-----QLAVEKHK 108

Db 69 ILLNGKPNNLNQEQIAKERRTELGFVFOFNLMPITVEENIILPLIDGKSVMKRQ 128

QY 109 WPEIPQVLHAGVLESFEEKPVFELSGEGEQORTALARVLKNPRIILADEPTGALDLTNS 168

Db 129 LAELSERL---GINHLKKRIAEISGGQQRVAVARAMTHHPOLLADEPTGALDLTNSK 185

QY 169 LVIEALRAL-ADKATVVVATHSPL 192

Db 186 DVNGLLQQLNEEBAATILMVTHDPL 210

RESULT 6

US-09-627-376-15

; Sequence 15, Application US/09627376

; Patent No. 6342385

; GENERAL INFORMATION:

; APPLICANT: Qi, Fengxia Caulfield, Page Chen, Ping

; TITLE OF INVENTION: MUTACIN I BIOSYNTHESIS GENES AND PROTEINS

; FILE REFERENCE: UAB-17402/22

; CURRENT APPLICATION NUMBER: US/09/627,376

; CURRENT FILING DATE: 2001-05-30

; NUMBER OF SEQ ID NOS: 17

; SOFTWARE: Patent in version 3.0

; SEQ ID NO 15

; LENGTH: 246

; TYPE: PRT

; ORGANISM: Streptococcus mutans

US-09-627-376-15

Query Match 31.6%; Score 322.5; DB 4; Length 246;

Best Local Similarity 35.8%; Pred. No. 7.4e-29;

Matches 78; Conservative 41; Mismatches 74; Indels 25; Gaps 4;

QY 2 IEINDLKKSFGVRLWQGLSHKFL-----PGTWTALTGASGSKSTLLNCLGTLDPK 53

Db 4 IGVSHLKKVYKQ---EGTNEALKDITFVSQGEFIAIMGSGSKSTLLNLLACMDYP 60

QY 54 SSGQILVEDDLLKLSRKORLYRKNVTGVLFQDYALIPDRTVKFNLQAV----- 104

Db 61 SSGHIIFNNTQLEKVKDEBAAVFRSRHIGFIFQFNULNLFNNKDNLLIPVIISGSKVNS 120

QY 105 -EKHKWPEIFQVLHAGVLESFEEKPVFELSGEGEQORTALARVLKNPRIILADEPTGALD 163

Db 121 YEK---RLDLAAVVGIESLLSKYPVELSGGQQRVAVARALIMNPDILADEPTGOLD 176

QY 164 LTRSELVIEALRALADKATVVVATHSPLPRESADTII 201

Db 177 SKTSORILNLLSNIAKRRTILMVTHSPKAAAYANRVL 214

RESULT 7

US-09-543-681A-5411

; Sequence 5411, Application US/09543681A

; Patent No. 6605709

; GENERAL INFORMATION:

; APPLICANT: GARY BRETON

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.1002-001

; CURRENT APPLICATION NUMBER: US/09/543,681A

; CURRENT FILING DATE: 2000-04-05

; PRIOR APPLICATION NUMBER: US 60/128,706

; PRIOR FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 8344

; SEQ ID NO 5411

; LENGTH: 653

; TYPE: PRT

; ORGANISM: Proteus mirabilis

US-09-543-681A-5411

Query Match 31.5%; Score 321; DB 4; Length 653;

Best Local Similarity 35.8%; Pred. No. 4.8e-28;

Matches 77; Conservative 46; Mismatches 78; Indels 14; Gaps 5;

QY 1 MIEINDLKKSFP---GVRILWQGLSHKFLPGTWTALTGASGSKSTLLNCLGTLDPKSSG 56

Db 10 LLELNVSRLYTNGEEDTVLVDISLTINAGEMVAITGASGSKSTLLNCLGTLDPKSSG 69

QY 57 QILVEDVDLLKLSRKORLYRKNVTGVLFQDYALIPDRTVKFNLQ-----AVEKHK 109

Db 70 EYKAGQCVADMESDQLAALRREHFGFIFQYHLMAHLTAEQNVIEIPAIYAGKSTEQRK- 128

QY 110 PEIPQVLHAGVL-ESFEEKPVFELSGEGEQORTALARVLKNPRIILADEPTGALDLTNS 168

Db 129 ERARALLTRGLAERIHVRP-SQLSGGQQRVAVARALMNGGEVILADEPTGALDSQSGK 187

QY 169 LVIEALRALADKATVVVATHSPLPRESADTIIKL 203

Db 188 EVMAILKQLNQOQHTVIVTHDPLIAQQAADRIIEI 222

RESULT 8

US-09-134-000C-6167

; Sequence 6167, Application US/09134000C

; Patent No. 6617156

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 032796-032

; CURRENT APPLICATION NUMBER: US/09/134,000C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/055,778

; PRIOR FILING DATE: 1997-08-15

; NUMBER OF SEQ ID NOS: 6812

; SOFTWARE: Patent in version 3.1

; SEQ ID NO 6167

; LENGTH: 239

; TYPE: PRT

; ORGANISM: Enterococcus faecalis

US-09-134-000C-6167

; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3822
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3832

Query Match 30.0%; Score 106; DB 4; Length 242;
Best Local Similarity 33.5%; Pred.No.5.8e-27;
Matches 73; Conservative 44; Mismatches 73; Indels 28; Gaps 4;

QY 1 MIEINDLKXKFGVRAILMQGLSHKFLPGMTALTGASGSGSKTLLNCLGTLDKPSSGGIIV 60
Db :
3 VINIKNLKKFGANLVRDINLTVEKEBVAIIGPSGSGSKTLRCMNLDDVPCKGVIP 62
QY 61 EDVDLLKLSTRKQRLRYKNVTGYLFQDYALLPDRTVKENLQLAVEKHKWPEIQVLHVG 120
Db :
63 EDNELTOHNVHLDLNRQK-MGMVFQNFPLPHKKVIENVMLA-----PLLHKDS 111
QY 121 LESPEEEXPVF-----ELSGGEQQOTALRVLLKNPRIILADEPTGALDL 164
Db :
112 KDQLKERALYLEKVKDKRADSPYNQLSGGQKQEVAIARALAMEPDVWLFEPTSALDP 171
QY 165 TNSBELVTEALRALADKGATVVVATHSPILF-RESADTII 201
Db :
172 EVVGDVLKVMRQLANESMTWIVITHENMFAKEISDKV 209

RESULT 13
US-09-107-532A-6667
; Sequence 6667, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FACIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012

%	TELECOMMUNICATION INFORMATION:
%	TELEPHONE: (781)893-5007
%	TELEFAX: (781)893-8277
%	INFORMATION FOR SEQ ID NO: 6667:
%	SEQUENCE CHARACTERISTICS:
%	LENGTH: 221 amino acids
%	TYPE: amino acid
%	TOPOLOGY: linear
%	MOLECULE TYPE: protein
%	HYPOTHETICAL: YES
%	ORIGINAL SOURCE:
%	ORGANISM: Enterococcus faecium
%	FEATURE:
%	NAME/KEY: misc_feature
%	LOCATION: (B) LOCATION 1...221
%	SEQUENCE DESCRIPTION: SEQ ID NO: 6667:
%	US-09-107-532A-6667
%	Query Match 30.0%; Score 305.5; DB 4; Length 221;
%	Best Local Similarity 37.5%; Pred. No. 5.8e-27;
%	Matches 78; Conservative 45; Mismatches 72; Indels 13; Gaps 6
Qy	1 MIEINDLKKSFG--VRILWGLSHKPLPGMTALTGASGSGKSTLLNCLGTLDKPSSQQ 57
Dd	6 ILLEKNVAYSANSKEVL-SGVNQKFELGKFYAVIGSKGTGKSTLLSLLAGLDRPOTCK 64
Qy	58 ILVEDVDLLKLSTRQRLYRNTVGYLFDYALIPDRTVKFNQLAVEKHKWEIPIQVLH 117
Dd	65 LIFKNED---IQNKGYSNHRKNISLFQNYYNLIDYLSPIENIRLNKGSAD----ESILF 117
Qy	118 AVGLESEFE-EKPVFELSGEQORTALARVLKKNPRIITLADEPTGALDLTNSSELVIAEIRA 176
Dd	118 ELGLDKKOIKRNVMKLSGGQQORVAIRALVSDAPIILADEPTGNLDSVTAGEIIINILKT 177
Qy	177 LA-DKGATVVVATHSPLPRESADIILK 203
Dd	178 LAKDRNKCIVIVVTHSKVEDADSADIILEL 205
%	RESULT 14
%	US-09-134-000C-6517
%	Sequence 6517, Application US/09134000C
%	Patent No. 6617156
%	GENERAL INFORMATION:
%	APPLICANT: Lynn Doucette-Stamm et al
%	TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
%	FILE REFERENCE: 032796-032
%	CURRENT APPLICATION NUMBER: US/09/134,000C
%	CURRENT FILING DATE: 1998-08-13
%	PRIOR APPLICATION NUMBER: US 60/055,778
%	PRIOR FILING DATE: 1997-08-15
%	NUMBER OF SEQ ID NOS: 6812
%	SOFTWARE: PatentIn version 3.1
%	SEQ ID NO 6517
%	Length: 221
%	Type: PRT
%	Organism: Enterococcus faecalis
%	US-09-134-000C-6517
%	Query Match 30.0%; Score 305.5; DB 4; Length 221;
%	Best Local Similarity 37.5%; Pred. No. 5.8e-27;
%	Matches 78; Conservative 45; Mismatches 72; Indels 13; Gaps 6
Qy	1 MIEINDLKKSFG--VRILWGLSHKPLPGMTALTGASGSGKSTLLNCLGTLDKPSSQQ 57
Dd	6 ILLEKNVAYSANSKEVL-SGVNQKFELGKFYAVIGSKGTGKSTLLSLLAGLDRPOTCK 64
Qy	58 ILVEDVDLLKLSTRQRLYRNTVGYLFDYALIPDRTVKFNQLAVEKHKWEIPIQVLH 117
Dd	65 LIFKNED---IQNKGYSNHRKNISLFQNYYNLIDYLSPIENIRLNKGSAD----ESILF 117
Qy	118 AVGLESEFE-EKPVFELSGEQORTALARVLKKNPRIITLADEPTGALDLTNSSELVIAEIRA 176

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 27, 2004, 16:36:31 ; Search time 45 Seconds
(without alignments)
1412.673 Million cell updates/sec

Title: US-09-868-338-9
Perfect score: 1019
Sequence: 1 MIEINDLKSGFGRILMOGL.....VVVATHSPFLRESADTIKL 203

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1288442 seqs, 313154207 residues

Total number of hits satisfying chosen parameters: 1288442

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published Applications AA:*
- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
 - 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
 - 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
 - 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
 - 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
 - 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
 - 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
 - 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
 - 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	399	39.2	213	16	US-10-474-776-389
2	380.5	37.3	210	16	US-10-474-776-386
3	371.5	36.5	248	14	US-10-156-761-11171
4	362.5	35.6	465	12	US-10-282-122A-64080
5	353	34.6	466	12	US-10-282-122A-63428
6	344.5	33.8	227	12	US-10-282-122A-67001
7	339.5	33.3	222	9	US-09-738-626-4456
8	339.5	33.3	222	12	US-10-627-476-334
9	339	33.3	648	9	US-09-815-242-13955
10	337	33.1	252	12	US-10-282-122A-44079
11	336.5	33.0	315	14	US-10-156-761-12964
12	336	33.0	236	9	US-09-815-243-13482
13	336	33.0	249	9	US-09-815-243-5642
14	336	33.0	252	9	US-09-815-242-12706
15	333.5	32.7	231	14	US-10-238-075-1256

16	332	32.6	252	12	US-10-282-122A-70852
17	330.5	32.4	663	9	US-09-815-242-11869
18	330.5	32.4	663	14	US-10-246-330-20
19	330	32.4	241	12	US-10-282-122A-60354
20	329.5	32.3	248	12	US-10-282-122A-63354
21	329.5	32.3	248	12	US-10-282-122A-64477
22	329.5	32.3	261	14	US-10-156-761-14933
23	328	32.2	252	12	US-10-282-122A-71920
24	325.5	31.9	247	15	US-10-369-493-10091
25	324.5	31.8	233	9	US-09-815-242-13827
26	324	31.8	248	12	US-10-282-122A-53217
27	324	31.8	257	9	US-09-815-242-10577
28	323.5	31.7	255	12	US-10-282-122A-61025
29	322.5	31.6	246	13	US-10-047-676A-15
30	321	31.5	271	9	US-09-815-242-13350
31	321	31.5	271	9	US-09-815-242-13690
32	320.5	31.5	256	12	US-10-282-122A-51558
33	315	30.9	242	12	US-10-282-122A-70384
34	314	30.8	240	12	US-10-282-122A-56935
35	314	30.8	379	12	US-10-282-122A-46319
36	312.5	30.7	243	12	US-10-282-122A-51603
37	312.5	30.7	262	14	US-10-156-761-11088
38	312	30.6	242	12	US-10-282-122A-70607
39	311	30.5	230	9	US-09-738-626-6814
40	311	30.5	230	12	US-10-627-476-334
41	310.5	30.5	245	12	US-10-282-122A-52619
42	310	30.4	222	12	US-10-282-122A-54068
43	310	30.4	227	12	US-10-282-122A-69667
44	309.5	30.4	228	9	US-09-815-242-10135
45	309.5	30.4	236	9	US-09-738-626-6304

ALIGNMENTS

RESULT 1
US-10-474-776-389
; Sequence 389, Application US/10474776
; Publication No. US00040110181A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; TITLE OF INVENTION: NOVEL STREPTOCOCCUS PNEUMONIAE OPEN READING FRAMES ENCODING POLYPEPTIDES
; FILE REFERENCE: AM100649-PCT
; CURRENT APPLICATION NUMBER: US/10/474,776
; CURRENT FILING DATE: 2003-10-14
; NUMBER OF SEQ ID NOS: 752
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 389
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-474-776-389

Query Match	39.2%;	Score	399;	DB	16;	Length	213;
Best Local Similarity	42.9%;	Pred. No.	3.3e-34;				
Matches	90;	Conservative	40;	Mismatches	70;	Indels	10;
Gaps	3;						
Qy	1	MIEINDLKSGFGRILMOGLSHKFLPCTMTALTGASGSGKSTLLNCILGTLDPKSSGOILV	60				
Db	1	MIDIQLEKFPNDRAIFSGNLKLERKGVYALTKSGSGKTTLLNLGKLEKIDGGRVLY	60				
Qy	61	EDVLLKLSRQRLYKNTVGVLFQDYALTPORTVFNQLQAVKHKWPEIPQVLHVG	120				
Db	61	QKDLKTIPTRE---YFRDQGVLYFNQGLLENQVSIKENLDLGFVGQKISKVERLERQVG	117				
Qy	121	LESFE-----EKVPFELSGGEOQTALARVLKNPRIILABPTGALDITNSLIVIA	173				
Db	118	ALAEKVNLYLDLEQKIYTLGSGEQRVALAKTIKNPLILADEPTAALDPENSEVMNL	177				
Qy	174	LRLADKATVAVATHSPFLRESADTIKL	203				
Db	178	LVDLKDNRILIIATHNPLVWKADEILDM	207				

```

RESULT 2
US-10-474-776-386
; Sequence 386, Application US/10474776
; Publication No. US20040110181A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; TITLE OF INVENTION: NOVEL STREPTOCOCCUS PNEUMONIAE OPEN READING FRAMES ENCODING POLYH
; FILE REFERENCE: AM100649-PCT
; CURRENT APPLICATION NUMBER: US/10/474,776
; CURRENT FILING DATE: 2003-10-14
; NUMBER OF SEQ ID NOS: 752
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 386
; LENGTH: 210
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-474-776-386

Query Match      37.3%; Score 380.5; DB 16; Length 210;
Best Local Similarity 41.0%; Pred. No. 3.1e-32;
Matches 86; Conservative 41; Mismatches 72; Indels 11; Gaps 4;

QY 1 MIEINDLKSPGVRILWQGLSHKFLPCTMTALTGASGSGKSTLLNCLGTLTKPSSGQILV 60
DB 1 MELKQVSKSPGERELFSNLMTPEAGKVVALIGSGSGKSTLLNMIGKLE-PYDGTIFY 59
QY 61 EDVLLKSLSTRKQRLKNTVGVLFQDYALIPDRTVKFNQLQAVEKHKWEI-----PQ 114
DB 60 RGKD---LANYKSDFFRHEGLYLFQNFGLIENOSIENKGLIGOKLSRSQRURQKQ 116
QY 115 VLHVGLESF-EKVPFELSGEQRTALARVLLKNPRIILADEPTGALDNLNSELVIEA 173
DB 117 ALEQVGLVLDLDRIFELSGESQRTALAKIILKNPFFILADEPTASIDPATSQLIMEI 176
QY 174 LRALADKATVVVATHSPPLFRESADTIKL 203
DB 177 LLSLRDNNRLIIATHNPAINEMADEVFTM 206

RESULT 3
US-10-156-761-11171
; Sequence 11171, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 11171
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-11171

Query Match      36.5%; Score 371.5; DB 14; Length 248;
Best Local Similarity 41.0%; Pred. No. 3.6e-31;
Matches 87; Conservative 37; Mismatches 77; Indels 11; Gaps 4;

QY 1 MIEINDLKSPGVRILWQGLSHKFLPCTMTALTGASGSGKSTLLNCLGTLTKPSSG 56
DB 9 IOLRSVSRRYGAGGAVTALDQ-VSLAFPRGTFTAVMGPSGSGKSTLLQCAAGLDRTSG 67
QY 57 QILVEDVLLKSLSTRKQRLKNTVGVLFQDYALIPDRTVKFN-----LQAVEKHKWEI 112
DB 68 SVTVGDTLTKLSETKTLTLRDRIGFVFOAFNLLPSLTAEONVALPRLAGRRPRTEV 127
QY 113 PQVLHVGLESF-EKVPFELSGEQRTALARVLLKNPRIILADEPTGALDNLNSELVIE 172
DB 128 REVLAQVGLGDRAGHRPTMGSGGQRTALARALITRDPVLFGDEPTGALDSQTSREVL 187
QY 173 ALRALAD-KGATVVVATHSPPLFRESADTIKL 203
DB 188 LLRGVMDSEGQTVIMVTHDPVAASYADRVVFL 219

RESULT 4
US-10-282-122A-64080
; Sequence 64080, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 64080
; LENGTH: 465
; TYPE: PRT
; ORGANISM: Mycoplasma pneumoniae
US-10-282-122A-64080

Query Match      35.6%; Score 362.5; DB 12; Length 465;
Best Local Similarity 36.7%; Pred. No. 8.1e-30;
Matches 77; Conservative 54; Mismatches 70; Indels 9; Gaps 4;

QY 1 MIEINDLKSPGVRILWQGLSHKFLPCTMTALTGASGSGKSTLLNCLGTLTKPSSG 56
DB 9 IOLRSVSRRYGAGGAVTALDQ-VSLAFPRGTFTAVMGPSGSGKSTLLQCAAGLDRTSG 67
QY 57 QILVEDVLLKSLSTRKQRLKNTVGVLFQDYALIPDRTVKFN-----LQAVEKHKWEI 112
DB 68 SVTVGDTLTKLSETKTLTLRDRIGFVFOAFNLLPSLTAEONVALPRLAGRRPRTEV 127
QY 113 PQVLHVGLESF-EKVPFELSGEQRTALARVLLKNPRIILADEPTGALDNLNSELVIE 172
DB 128 REVLAQVGLGDRAGHRPTMGSGGQRTALARALITRDPVLFGDEPTGALDSQTSREVL 187
QY 173 ALRALAD-KGATVVVATHSPPLFRESADTIKL 203
DB 188 LLRGVMDSEGQTVIMVTHDPVAASYADRVVFL 219

```


Qy	81	YQYLFDQVALIPDRTVKFNLCIAVEKHWP-----ETPOVLHVAGLSFPSEFKPVFELSQGE	136
Db	316	YGFYFQQYGLFNLTVRENBIGANLOPDFSKRISDALLEAVGMDLQKQUPNELSGQ	375
Qy	137	QORTALARVLKNPRITILADEPTGALDTNSELVIEALRALADK-GATVVVATHSPFPRE	195
Db	376	QQRVSARAFAKNPILIFGDEFTGALDLEMTQVLKQFLAKKRYQYTWIIVTHNNLIAN	435
Qy	196	SADTII 201	
Db	436	LADLVI 441	

```

US-10-282-122A-8/001
; Sequence 67001, Application US/10282122A
; Publication NO. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malore, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A

```

```

, CURRENT FILING DATE: 2003-02-20
, PRIOR APPLICATION NUMBER: 60/191,078
, PRIOR FILING DATE: 2000-03-21
, PRIOR APPLICATION NUMBER: 60/206,848
, PRIOR FILING DATE: 2000-05-23
, PRIOR APPLICATION NUMBER: 60/207,727
, PRIOR FILING DATE: 2000-05-26
, PRIOR APPLICATION NUMBER: 60/230,335
, PRIOR FILING DATE: 2000-09-06
, PRIOR APPLICATION NUMBER: 60/230,347
, PRIOR FILING DATE: 2000-09-09
, PRIOR APPLICATION NUMBER: 60/242,578
, PRIOR FILING DATE: 2000-10-23
, PRIOR APPLICATION NUMBER: 60/253,625
, PRIOR FILING DATE: 2000-11-27
, PRIOR APPLICATION NUMBER: 60/257,931
, PRIOR FILING DATE: 2000-12-22
, PRIOR APPLICATION NUMBER: 60/267,636
, PRIOR FILING DATE: 2001-02-09
, PRIOR APPLICATION NUMBER: 60/269,308
, PRIOR FILING DATE: 2001-02-16
, Remaining Prior Application data removed - See File Wrapper or PALM.
, NUMBER OF SEQ ID NOS: 78614
, SOFTWARE: PatentIn version 3.1
, SEQ ID NO 67001
, LENGTH: 227
, TYPE: PRT
, ORGANISM: Pasteurella multocida
,
,
, Query Match 33.8%; Score 344.5; DB 12; Length 227;
US-10-282-122A-67001

```

[illegible]

Db 61 DGVDAQDLBGRFRFAEKIGLVFQFHLIPLYTALENVMLAQHYSHVIDEAAKAVLIQ 120
 Qy 118 AVGL-ESFEKVPFELSGGQORTALARVLLKNPRIILADEPTGALDNLNSELVIEA 176
 Db 121 QVGLAHFEDRP-SQLSGGQQRVCIAIALVNPFPVIFADEPTGNLDKNEQLVLDLVT 179
 Qy 177 LADKATVVVATHSPFPRESADTIKL 203
 Db 180 LNOQRTVVMVTHNPESKLADRTIFL 206

RESULT 7

US-09-738-626-4456
 ; Sequence 4456, Application US/09738626
 ; Publication No. US20020197605A1
 ; GENERAL INFORMATION:
 ; APPLICANT: NAKAGAWA, SATOSHI
 ; APPLICANT: MIZOGUCHI, HIROSHI
 ; APPLICANT: ANDO, SEIKO
 ; APPLICANT: HAYASHI, MIKIRO
 ; APPLICANT: OCHIALI, KEIKO
 ; APPLICANT: YOKOI, HARUHIKO
 ; APPLICANT: TATEISHI, NAKO
 ; APPLICANT: SENOH, AKIHIKO
 ; APPLICANT: IKEDA, MASATO
 ; APPLICANT: OZAKI, AKIO
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; CURRENT APPLICATION NUMBER: US/09/738,626
 ; CURRENT FILING DATE: 2000-12-18
 ; PRIOR APPLICATION NUMBER: JP 99/377484
 ; PRIOR FILING DATE: 1999-12-16
 ; PRIOR APPLICATION NUMBER: JP 00/159162
 ; PRIOR FILING DATE: 2000-04-07
 ; PRIOR APPLICATION NUMBER: JP 00/280988
 ; PRIOR FILING DATE: 2000-08-03
 ; NUMBER OF SEQ ID NOS: 7059
 ; SOFTWARE: Patent in ver. 3.0
 ; SEQ ID NO 4456
 ; LENGTH: 222
 ; TYPE: PRT
 ; ORGANISM: Corynebacterium glutamicum
 US-09-738-626-4456

Query Match 33.3%; Score 339.5; DB 9; Length 222;
 Best Local Similarity 38.1%; Pred. No. 8.3e-28;
 Matches 80; Conservative 36; Mismatches 83; Indels 9; Gaps 3;
 Qy 2 IEINDLKKSPG-----VRILWQGLSHKFLPQTMTALTGASGSGKSTLLNCLGTLDPKSSG 56
 Db 8 LELQNSICAFGEGRPHVSAL-NNVSLAVNFGELVAMGPGSGKSTLLNCLGTLDPKSSG 66
 Qy 57 QILVEDVDLLKSTRKQRLYKNTVGYLFODYALIPDRTVKFNQLAVE---KHKWPEIP 113
 Db 67 HVLIDGASADLNKRAAETRRRHIGVIFQNYNLVPTLTGENVGLPFLDGTDRQAVA 126
 Qy 114 QVLHAGVLESFEEKVPFELSGGQORTALARVLLKNPRIILADEPTGALDNLNSELVIEA 173
 Db 127 IALAEVGLGEGFDDRPPEISGGQQRVAIARALIGPRKILLADEPTGALDNLNSELVIEA 186
 Qy 174 LRALADKATVVVATHSPFPRESADTIKL 203
 Db 187 LRQRIDSGAAGLLVTHPRFAAWADRTIML 216

RESULT 8

US-10-627-476-442
 ; Sequence 442, Application US/10627476
 ; Publication No. US20040030116A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Pompejus, Mark
 ; APPLICANT: Kroger, Burkhard
 ; APPLICANT: Schoder, Hartwig

; APPLICANT: Zelder, Oskar
 ; APPLICANT: Haberhauer, Gregor
 ; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS
 ; TITLE OF INVENTION: INVOLVED IN MEMBRANE SYNTHESIS AND MEMBRANE
 ; FILE REFERENCE: BGI-125CPCN
 ; CURRENT APPLICATION NUMBER: US/10/627,476
 ; CURRENT FILING DATE: 2003-07-25
 ; PRIOR APPLICATION NUMBER: 09/602,787
 ; PRIOR FILING DATE: 2000-06-23
 ; PRIOR APPLICATION NUMBER: USSN 60/141031
 ; PRIOR FILING DATE: 1999-06-25
 ; PRIOR APPLICATION NUMBER: DE 19931454.3
 ; PRIOR FILING DATE: 1999-07-08
 ; PRIOR APPLICATION NUMBER: DE 19931478.0
 ; PRIOR FILING DATE: 1999-07-08
 ; PRIOR APPLICATION NUMBER: DE 19931563.9
 ; PRIOR FILING DATE: 1999-07-08
 ; PRIOR APPLICATION NUMBER: DE 19932122.1
 ; PRIOR FILING DATE: 1999-07-09
 ; PRIOR APPLICATION NUMBER: DE 19932124.8
 ; PRIOR FILING DATE: 1999-07-09
 ; PRIOR APPLICATION NUMBER: DE 19932125.6
 ; PRIOR FILING DATE: 1999-07-09
 ; PRIOR APPLICATION NUMBER: DE 19932128.0
 ; PRIOR FILING DATE: 1999-07-09
 ; PRIOR APPLICATION NUMBER: DE 19932180.9
 ; PRIOR FILING DATE: 1999-07-09
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 678
 ; SEQ ID NO 442
 ; LENGTH: 222
 ; TYPE: PRT
 ; ORGANISM: Corynebacterium glutamicum
 US-10-627-476-442

Query Match 33.3%; Score 339.5; DB 12; Length 222;
 Best Local Similarity 38.1%; Pred. No. 8.3e-28;
 Matches 80; Conservative 38; Mismatches 83; Indels 9; Gaps 3;
 Qy 2 IEINDLKKSPG-----VRILWQGLSHKFLPQTMTALTGASGSGKSTLLNCLGTLDPKSSG 56
 Db 8 LELQNSICAFGEGRPHVSAL-NNVSLAVNFGELVAMGPGSGKSTLLNCLGTLDPKSSG 66
 Qy 57 QILVEDVDLLKSTRKQRLYKNTVGYLFODYALIPDRTVKFNQLAVE---KHKWPEIP 113
 Db 67 HVLIDGASADLNKRAAETRRRHIGVIFQNYNLVPTLTGENVGLPFLDGTDRQAVA 126
 Qy 114 QVLHAGVLESFEEKVPFELSGGQORTALARVLLKNPRIILADEPTGALDNLNSELVIEA 173
 Db 127 IALAEVGLGEGFDDRPPEISGGQQRVAIARALIGPRKILLADEPTGALDNLNSELVIEA 186
 Qy 174 LRALADKATVVVATHSPFPRESADTIKL 203
 Db 187 LRQRIDSGAAGLLVTHPRFAAWADRTIML 216

RESULT 9

US-09-815-242-13995
 ; Sequence 1395, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; TITLE OF INVENTION: Prokaryotes
 ; FILE REFERENCE: ELITRA.011A

```
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13995
; LENGTH: 648
; TYPE: PRT
; ORGANISM: Salmonella typhi
US-09-815-242-13995
```

```
Query Match 33.3%; Score 339; DB 9; Length 648;
Best Local Similarity 36.9%; Pred. No. 4.3e-27;
Matches 79; Conservative 42; Mismatches 81; Indels 12; Gaps 3;

QY 1 MIEINDLKKSF-----GVRILWQGLSHKFLPTMTALTGASGSKSTLLNCLGTLDPKPS 55
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4 LLELCNVSRSYPSGSEQAVNL-KDLSQHAGENVAIVGVSGSKSTLMMILGCLDKPTS 62
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 56 GQILVEDVDLLKSTRKQRLRYKNTVGYLFQDYALIPDRIVKFNQLAV-----EKHW 109
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 63 GYRVAGRDVSTLDPDLAAQLREHFHFIFORYHLLSHLTAAQNVPEIPYVAGIERKQ 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 110 PEIPQVLHAGVLESPEEKVPFELSGEQOORTALARVLLKNPRIILADEPTGALDITNSEL 169
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 123 ABARELLRLGLSDRDYDPPSOLSGGQQOORVSIARALMNGGVILADEPTGALDSHGER 182
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 170 VTEALRALADKATVVVATHSPFPRESADTIKL 203
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 183 VMAILRQLRDRGHTVIIVTHDPLIAAQAERIIEI 216
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
RESULT 10
US-10-282-122A-44079
; Sequence 44079, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Cart, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
```

```
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 44079
; LENGTH: 252
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-282-122A-44079
```

```
Query Match 33.1%; Score 337; DB 12; Length 252;
Best Local Similarity 34.7%; Pred. No. 1.8e-27;
Matches 74; Conservative 47; Mismatches 82; Indels 10; Gaps 3;

QY 1 MIEINDLKKSPGVRILWQGL---SHKELPGMTALTGASGSKSTLLNCLGTLDPKPS 57
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 ILSVQHVSKTYGKHTTQALKDINFDIQKGEFVAIMGPSGSKTLLNVLSSIDQISSGS 62
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 58 ILVEDVDLLKSTRKQRLRYKNTVGYLFQDYALIPDRIVKFNQLAVKHKWPEIQ--- 114
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 63 VIANGQELNKLNQALAKFRKESIGTFQDYSLPILTVKENIMPLSVQKSKATMEEN 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 115 ---VLHAGVLESPEEKVPFELSGEQOORTALARVLLKNPRIILADEPTGALDITNSELVI 171
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 123 YKAITTALGIYDLGNKYPSELSGGQQOORTAARAFVHKPQIIFADEPTGALDSKANDLL 182
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 172 BALRALADK-GATVVVATHSPFPRESADTIKL 203
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 183 QRLEEMKNSFTTIVVTHDPAASFAERVIML 215
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
RESULT 11
US-10-156-761-12964
; Sequence 12964, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-282
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12964
; LENGTH: 315
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-12964
```

```
Query Match 33.0%; Score 336.5; DB 14; Length 315;
Best Local Similarity 37.7%; Pred. No. 2.8e-27;
Matches 80; Conservative 39; Mismatches 82; Indels 11; Gaps 4;
```



```
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12706
; LENGTH: 252
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12706

Query Match      33.8%; Score 336; DB 9; Length 252;
Best Local Similarity 34.7%; Pred. No. 2.3e-27;
Matches 74; Conservative 47; Mismatches 82; Indels 10; Gaps 3;

QY 1 MIEINDLKXSGFVRILWQGL---SHKELPGTMTALTGASGSGKSTLLNCLGTLDPKPSGQ 57
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 ILSVQHVSKTYGKXHTFQALKDNFDIQKGEFVALMGPSGSKTLLNVLSSIDQISSGS 62
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 58 ILVEDVDLLKSLRKQRLRYNTVGYLFQDYALIPDRTVFNQLAVEKHKWEIPO--- 114
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 63 VIANGQELNKLNOAKAKFKRESIGFTFDYSILPTLVKXENIMLPLSVQKMSKATMEKN 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 115 ---VLHAVGLSPKPEKVPFELSGEQOQTALARVLLKNPRIILADEPTGALDITNSELVI 171
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 123 YKAITTALGIYDLGNKYPSELSGQQOQTAAARAFVHKPQIIFADEPTGALDSKANDLL 182
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 172 EALRALADK-GATVVVATHSPLPRESADTIKL 203
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 183 QRLEENKSFDTITVMVTHDPVAASFAERVIML 215
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
RESULT 15
US-10-238-075-1256
; Sequence 1256, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
; APPLICANT: I.N.S.E.R.M.
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolated
; TITLE OF INVENTION: E.coli, and biological uses of these polynucleotides and of their
; FILE REFERENCE: BLANDINE
; CURRENT APPLICATION NUMBER: US/10/238,075
; CURRENT FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1256
; LENGTH: 231
; TYPE: PRT
```

```
; ORGANISM: Escherichia coli
US-10-238-075-1256

Query Match      32.7%; Score 333.5; DB 14; Length 231;
Best Local Similarity 38.0%; Pred. No. 3.6e-27;
Matches 78; Conservative 32; Mismatches 92; Indels 3; Gaps 1;

QY 2 IEINDLKXSGFVRILWQGLSHKFLPGTMTALTGASGSGKSTLLNCLGTLDPKPSGQILVE 61
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 11 IETRHLYKRFQDVTALDDINLRIAQCEFAVIMGASGSGKTTLMNLTICLTATGEGQVFLD 70
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 62 DVDLLKSLTRKORLYRKNTVGYLFQDYALIPDRTVKNLQLAVERKHKWE---IPQVLHA 118
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 71 GTDAALDEGRRRRAEKIGLVFOQFHLIPTALENIMLAQHYASVVDEAAARKVLEQ 130
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 119 VGLESFEEKVPFELSGEQOQTALARVLLKNPRIILADEPTGALDITNSELVIEALRALA 178
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 131 VGLGHRVTHLPQSLSGEQOQRCVCIARALVNEPPVIFADEPTGNLDEENQORVLDLTLH 190
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 179 DKGATVVVATHSPLPRESADTIKL 203
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 191 ROGRITVMVTHNPALGQFADRIILRL 215
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Search completed: July 27, 2004, 16:41:55
Job time : 47 secs

THIS PAGE BLANK (USPTO)

DB 178 LVLDKDNRIIIIIATHNPLVWVKADEIIDM 207

RESULT 4

G98096

hypothetical protein ABC-NBD [imported] - Streptococcus pneumoniae (strain R6)

C:Species: Streptococcus pneumoniae

C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001

C:Accession: G98096

C:Residues: 1-213 <KUR>

A:Cross-references: GB:AE007317; PIDN:AAK99406.1; PID:gl5458184; GSPDB:GN00174

C:Genetics:

A:Gene: ABC-NBD

Query Match 39.7%; Score 405; DB 2; Length 213;

Best Local Similarity 41.9%; Pred. No. 7.9e-28;

Matches 88; Conservative 43; Mismatches 69; Indels 10; Gaps 3;

QY 1 MIEINDLKXFGVIRILWQGLSHKFLPGTMTALTGASGSGKSTILNCLGTLDPKSSGQILV 60

DB 1 MIEINDLKXFGVIRILWQGLSHKFLPGTMTALTGASGSGKSTILNCLGTLDPKSSGQILV 60

QY 61 EDVLLKLSRKQRLYRKNTVGYLFQDYALIPDRVFNQLAV-----EKHKWPEIPQ 114

DB 61 DGTSLKDI---KPSVFRDYLGYLFQDFGLIESQTVKENLNLGLVGGKLEKEKISLMKQ 117

QY 115 VLHVGLESFB-EKPVFELSGEQORTALARVLKNPRIILADEPTGALDITNSSELVIEA 173

DB 118 ALNRVNLISYLDLQKPIFELSGEQORTALARVLKNPRIILADEPTGALDITNSSELVIEA 177

QY 174 LRALADKATVVVATHSPRESADTIKL 203

DB 178 LESLKNPRTIIIIATHNPLVWVKADEIIDM 207

RESULT 5

E95232

ABC transporter, ATP-binding protein SP1987 [imported] - Streptococcus pneumoniae (strain R6)

C:Species: Streptococcus pneumoniae

C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001

C:Accession: E95232

C:Residues: 1-213 <KUR>

A:Cross-references: GB:AE005672; PIDN:AAK76054.1; PID:gl4973495; GSPDB:GN00164; TIGR:SP4

A:Experimental source: strain TIGR4

C:Genetics:

A:Gene: SP1987

Query Match 39.2%; Score 399; DB 2; Length 213;

Best Local Similarity 42.9%; Pred. No. 2.6e-27;

Matches 90; Conservative 40; Mismatches 70; Indels 10; Gaps 3;

QY 1 MIEINDLKXFGVIRILWQGLSHKFLPGTMTALTGASGSGKSTILNCLGTLDPKSSGQILV 60

DB 1 MIEINDLKXFGVIRILWQGLSHKFLPGTMTALTGASGSGKSTILNCLGTLDPKSSGQILV 60

QY 61 EDVLLKLSRKQRLYRKNTVGYLFQDYALIPDRVFNQLAV-----EKHKWPEIPQ 120

DB 61 QKDKLTIPTRE---YFRDQGYLFQDFGLIENSIKENLDLGFVGQKISKVERLERQVG 117

QY 121 -LESFE-----EKPVFELSGEQORTALARVLKNPRIILADEPTGALDITNSSELVIEA 173

DB 118 ALEKVNGLVLDLEQKIYTLGSGEQORTALARVLKNPRIILADEPTGALDITNSSELVIEA 177

QY 174 LRALADKATVVVATHSPRESADTIKL 203

DB 178 LESLKNPRTIIIIATHNPLVWVKADEIIDM 207

RESULT 6

E95013

hypothetical protein SP0111 [imported] - Streptococcus pneumoniae (strain TIGR4)

C:Species: Streptococcus pneumoniae

C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001

C:Accession: A95013

C:Residues: 1-213 <KUR>

A:Cross-references: GB:AE005672; PIDN:AAK74298.1; PID:gl4971578; GSPDB:GN00164; TIGR:SP4

A:Experimental source: strain TIGR4

C:Genetics:

A:Gene: SP0111

Query Match 37.9%; Score 386; DB 2; Length 213;

Best Local Similarity 40.2%; Pred. No. 3.6e-26;

Matches 84; Conservative 45; Mismatches 74; Indels 6; Gaps 1;

QY 1 MIEINDLKXFGVIRILWQGLSHKFLPGTMTALTGASGSGKSTILNCLGTLDPKSSGQILV 60

DB 178 LVLDKDNRIIIIIATHNPLVWVKADEIIDM 207

RESULT 4

G98096

hypothetical protein ABC-NBD [imported] - Streptococcus pneumoniae (strain R6)

C:Species: Streptococcus pneumoniae

C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001

C:Accession: G98096

C:Residues: 1-213 <KUR>

A:Cross-references: GB:AE007317; PIDN:AAK99406.1; PID:gl5458184; GSPDB:GN00174

C:Genetics:

A:Gene: ABC-NBD

Query Match 39.2%; Score 399; DB 2; Length 213;

Best Local Similarity 42.9%; Pred. No. 2.6e-27;

Matches 90; Conservative 40; Mismatches 70; Indels 10; Gaps 3;

QY 1 MIEINDLKXFGVIRILWQGLSHKFLPGTMTALTGASGSGKSTILNCLGTLDPKSSGQILV 60

DB 1 MIEINDLKXFGVIRILWQGLSHKFLPGTMTALTGASGSGKSTILNCLGTLDPKSSGQILV 60

QY 61 EDVLLKLSRKQRLYRKNTVGYLFQDYALIPDRVFNQLAV-----EKHKWPEIPQ 120

DB 61 QKDKLTIPTRE---YFRDQGYLFQDFGLIENSIKENLDLGFVGQKISKVERLERQVG 117

QY 121 -LESFE-----EKPVFELSGEQORTALARVLKNPRIILADEPTGALDITNSSELVIEA 173

DB 118 ALEKVNGLVLDLEQKIYTLGSGEQORTALARVLKNPRIILADEPTGALDITNSSELVIEA 177

QY 174 LRALADKATVVVATHSPRESADTIKL 203

DB 178 LVLDKDNRIIIIIATHNPLVWVKADEIIDM 207

RESULT 5

E95013

hypothetical protein SP0111 [imported] - Streptococcus pneumoniae (strain TIGR4)

C:Species: Streptococcus pneumoniae

C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001

C:Accession: A95013

C:Residues: 1-213 <KUR>

A:Cross-references: GB:AE005672; PIDN:AAK74298.1; PID:gl4971578; GSPDB:GN00164; TIGR:SP4

A:Experimental source: strain TIGR4

C:Genetics:

A:Gene: SP0111

Query Match 37.9%; Score 386; DB 2; Length 213;

Best Local Similarity 40.2%; Pred. No. 3.6e-26;

Matches 84; Conservative 45; Mismatches 74; Indels 6; Gaps 1;

QY 1 MIEINDLKXFGVIRILWQGLSHKFLPGTMTALTGASGSGKSTILNCLGTLDPKSSGQILV 60

Db 1 MIELKNIITKTGGKIVILONLSRIDQGLVAIVKSGSGKSTLLNLLGLIDCDYSGRYEI 60

QY 61 EDVLLKLSRQRKYRNTVGYIFODVALPDRVTXKENIQLAV-----EKHKWPBPQ 114

Db 61 FQOTNLVANSKSOPIIIEHYSYIFQNFALIDDETVYNLMALKYVKPKDKLKVVEE 120

QY 115 VLVHVGLSFBEKPVFELSGEQOQTALARVLLKNPRIILADEPTGALDLTNSLVTEAL 174

Db 121 ILERVGLSATLHQRVSELSGGEQRIAVARAILFSPQLILADEPTGSDPENRDLVXKFL 180

QY 175 RALADKGATVVVAHSPLFRESADTIIKL 203

Db 181 LEMNREGKTVIIVTHDAYVAQCCHRIEEL 209

RESULT 6

D37894
 Hypothetical protein ABC-NBD [imported] - Streptococcus pneumoniae (strain R6)
 C:Species: Streptococcus pneumoniae
 C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
 C:Accession: D37894
 e; Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgess, S.; DeHoff, B.S.; E
 r, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
 Y, P.; Sun, P.M.; Winkler, M.E.
 J. Bacteriol. 183, 5709-5717, 2001
 A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
 A>Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
 A:Reference number: A97872; MUID:21423245; PMID:11544234
 A:Accession: D37894
 A>Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-213 <KUR>
 A:Cross-references: GS:AE007317; PIDN:AAK98904.1; PID:g15457636; GSPDB:GN00174
 C:Genetics:
 A:Gene: ABC-NBD

Query Match	37.8%	Score	385;	DB 2;	Length	213;			
Best Local Similarity	39.7%	Pred. No.	4.3-26;						
Matches	83;	Conservative	46;	Mismatches	74;	Indels	6;	Gaps	1;
Qy	1	MIE	NDLKKSGFVRIILWQGLSHKFLPTMTALTGASGSGKSTLNCILGTLDKPSGGQILV	60					
Db	1	MIE	LNKTKITGGKVIIDNLRLDQGLVAIVGSGSGKSTLNLGLIDGQYSGRYEI	60					
Qy	61	EDV	LLKLSTRKQRLYRNTVGYLFQDYALIFDRTVFNLQLAV-----EKHKWEIPIQ	114					
Db	61	FGQ	TNLVAVNSAKSQTIIREHISYLFQNFALIDDETVENLMALKYVKLPKKOKLKKVEE	120					
Qy	115	VLA	HVGELSEEEKVFEVLSGGEQRTALARVLLKNPRLILADETGAIDLTNSELVTEAL	174					
Db	121	ILERV	GLSATLHQRVSELSGGEQRIAVARAILKPSQILILADETGLSDPENRDLVLKFL	180					
Qy	175	RALAD	KATVVVAVTHSPFLFRESADTIIKL	203					
Db	181	LEMM	REGKTVIIVTHDAYVAOCHRVTEL	209					

RESULT 7

G95228
ABC transporter, ATP-binding protein SPI957 [imported] - Streptococcus pneumoniae (strain G95228)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: G95228
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
nson, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
nson, T.; Hickney, E.A.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A.; Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: G95228
A:Status: preliminary
A:Molecule type: DNA

A;Residues: 1-210 <KUR>
A;Cross-references: GB:
A;Experimental source:
C;Genetics:
A;Gene: SP1957

```

Query Match.          37.3%; Score 380.5; DB 2; Length 210;
Best Local Similarity 41.0%; Pred. No. 1e-25;
Matches 86; Conservative 41; Mismatches 72; Indels 11; Gaps 4;

QY 1 MIEINDKKSGFGRILWQGLSHKFLPTMTALTGASGSKTILNCLGTLDKSSGQILV 60
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 MIELKQVSKSGFGRFLFNSLWTFEAGKVYALGSSGSGKTTLNWIKGLE-PYDGFIFY 59
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 61 EDVDLLKLSRTRKORLYRNTGVLYFDYALIPDRTVKFNQLQAVEKHKWEI-----PQ 114
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 60 RGRD---LANYKSSDFRHELGYLFONFGLIENQSIENLKLGLIGQKLSRSEQLRQK 116
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 115 VLHAVGLGSEF-EKVPFELSGGQQRATLARVLKKNPRIILADEPTGALDITNSLVIEA 173
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 117 ALFQVGLVLDLXRFELSGGESQSRVALAKILKNPPFLADEPTASIDPATSLIMEI 176
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 174 LRLAKGATVTVVATHSPFRESADTIKL 203
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 177 LLSLRDNDRLIIATHNPAINWEMADVFVFM 206
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 8

C98093
hypothetical protein ABC-NBD [imported] - Streptococcus pneumoniae (strain R6)
C/Species: Streptococcus pneumoniae
C/Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C/Accession: C98093
R/Roskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff
e, K.; Leblanc, D.J.; Lee, I.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAb
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A/Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jasku
A/Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A/Reference number: A97872; MUID:21429245; PMID:11544234
A/Accession: C98093
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-210 <KUR>
A/Cross-references: GB:A8007317; PIDN:AAL00576.1; PID:gl5459455; GSPDB:GN00174
A/Genetic: C
A/Gen: ABC-NBD

Query Match	37.3%	Score 380.5;	DB 2;	Length 210;
Best Local Similarity	41.0%;	Pred. No. 1e-25;		
Matches 86;	Conservative 41;	Mismatches 72;	Indels 11;	Gaps 4;

QY	1	MIEINDLKSGVIRLWQGS	SHKFLPCTMTALT	CASGSGKSTLLN	CLGTIDKPSQILV	60		
		:	:	:	:			
Db	1	MIELQVSKSGERELFSL	SMTFEAGKVALL	GSSGKTLIMN	IGKLEPYDTIFY	59		
		:	:	:	:			
QY	61	EDVLLKLTSTKRLYK	NTVGYLFQDYAL	IPRTKFNQLAVE	KHKEI-----	114		
		:	:	:	:			
Db	60	RKGO---	LANYKSDFFR	HSLGYLFQNG	FIENQSIENKLG	IGCKLSRSEORLRQ	116	
		:	:	:	:	:		
QY	115	VLHVALESFE	-EXPVFELSGG	EQORTALARVLL	KNPRIILADE	PTGALDLTNSL	VEIFA	173
		:	:	:	:	:	:	
Db	117	ALEQVGLVYLD	KRIEELSGGS	QORVALAKIIL	KNPFIILADE	PTASIDPAT	SQLINEI	176
		:	:	:	:	:	:	
QY	174	LRALADKGA	TVVWATHSPL	FRESADTIKL				203
		:	:	:	:	:	:	
Db	177	LLSRDRNR	LIITIIATHN	PAIWMAD	VEFTM			206
		:	:	:	:	:	:	

RESULT 9

A84058
ABC transporter (ATP-binding protein) BH3505 [imported] - *Bacillus halodurans* (strain C-3002)
C/Species: *Bacillus halodurans*

NUMERIC ACTUS RES. 26, 4317-4331, 2000

QY 170 VIEALRALAD-KGATVVVATHSPLEFRESADTIKL 203

MAILING LABELS REQ: 20, 101/1034, 2000

Db 182 VMEIRHONVGLVTIILVTHDPSLAKYGRVIRL 216

Search completed: July 27, 2004, 16:36:52
Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 27, 2004, 16:29:29 ; Search time 13 Seconds

(without alignments)
813.095 Million cell updates/sec

Title: US-09-868-338-9

Perfect score: 1019

Sequence: 1 MEINDLKKSFGVRIILWQGL.....VVVATHSPILFRSADTIKL 203

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	362.5	35.6	465	1 Y065 MYCPN	P75612 mycoplasma
2	353	34.6	466	1 Y065 MYCGE	P47311 mycoplasma
3	347	34.1	648	1 MACB ECOLI	P75831 escherichia
4	337	33.1	257	1 YXDL BACSU	P42423 bacillus su
5	329.5	32.3	248	1 Y986 MYCTU	O53899 mycobacteri
6	325.5	31.9	284	1 Y46B MYCGE	O92b70 mycoplasma
7	325	31.9	240	1 YQIZ BACSU	P54537 bacillus su
8	323	31.7	234	1 Y352 THEMA	O9wy17 thermotoga
9	312.5	30.7	284	1 Y46B MYCPN	Q50316 mycoplasma
10	311.5	30.6	235	1 Y796 METJA	Q58206 methanococc
11	309.5	30.4	233	1 LOLE ECOLI	P75957 escherichia
12	309	30.3	224	1 YF08 METJA	Q58903 methanococc
13	309	30.3	344	1 METN VIBCH	O9ktj5 vibrio chol
14	303	29.7	247	1 YCKI BACSU	P39456 bacillus su
15	302.5	29.7	249	1 YXEO BACSU	P54954 bacillus su
16	301	29.5	330	1 YP64 MYCTU	Q50734 mycobacteri
17	300	29.4	377	1 POTG ECOLI	P31134 escherichia
18	296	29.0	250	1 YECC ECOLI	P37774 escherichia
19	295	28.9	345	1 METN HAEIN	P44785 haemophilus
20	291	28.6	222	1 FTSE ECOLI	P10115 escherichia
21	289.5	28.4	231	1 YTRE BACSU	O34392 bacillus su
22	289	28.4	343	1 METN YERPE	Q8zh38 yersinia pe
23	287.5	28.2	221	1 Y700 RICPR	Q9zcm4 rickettsia
24	287.5	28.2	229	1 LOLE BUCAP	Q44613 buchnera ap
25	286	28.1	231	1 LOLE NEIMP	P57031 neisseria m
26	285	28.0	343	1 METN ECOL6	O8x729 escherichia
27	285	28.0	343	1 METN ECOLI	P30750 escherichia
28	282.5	27.7	231	1 LOLE NEIMA	P57030 neisseria m
29	282.5	27.7	348	1 CYSA_XILPT	O87ct9 xylella fas
30	282	27.7	355	1 CYSA SYN33	P74548 synecocyst
31	280.5	27.5	339	1 Y467 MYCPN	P75110 mycoplasma
32	280	27.5	242	1 GLNQ BACST	P27675 bacillus st
33	279.5	27.4	227	1 LOLE_HAEIN	P45247 haemophilus

ALIGNMENTS

```

RESULT 1
Y065 MYCPN
ID Y065 MYCPN STANDARD; PRT; 465 AA.
AC P75612;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical ABC transporter ATP-binding protein MG065 homolog
DE (R02 orf465).
DE
GN MPN081 OR MP074.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R., Hilbert H., Pragens H., Pirkl E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae."
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -!- SIMILARITY: Belongs to the ABC transporter family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AB000009; AAB95722.1; -
CC PIR; S73400; S73400.
CC InterPro; IPR003593; AAA_ATPase.
CC InterPro; IPR003439; ABC_transporter.
CC Pfam; PF00005; ABC_tran; 1.
CC ProDom; PD000006; ABC_transporter; 1.
CC SMART; SM00382; AAA; 1.
CC PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
CC PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
CC KW Hypothetical protein; ATP-binding; Transport; Complete proteome.
CC NP_BIND 268 275 ATP (POTENTIAL).
CC SQ SEQUENCE 465 AA; 53511 MW; P8F85014C4A1A25D CRC64;

```

```

Query Match 35.6%; Score 362.5; DB 1; Length 465;
Best Local Similarity 36.7%; Pred. No. 3.1e-23;
Matches 77; Conservative 54; Mismatches 70; Indels 9; Gaps 4;
QY 1 MEINDLK--SFGV--RIILWQGLSHKFLPCTMTALTGASGSGKSTILNCIGLTKDPSSG 56
Db 231 IIEUNWIKYITNGVTTNAVNLKGLDKLKAHDFVILGPSGSGKATLLNIIISGMDRPSG 290
QY 57 QILVEDVLLKLSTRKQRLRYKNTVGYLFQDYALIPDTRVKFNQLQAVEKHKPE---I 112
Db 291 SVVWNGQEMICMNDQLNTFRNRYVGVYFQYGLPLNLTVRNVEVNGANLQRPDKRINI 350

```

```

Qy 113 POVLHAYGLSEFEKPKVPELSSGSGQORTALARVLLKNPRLIILADEPTGALDNLSELVIE 172
Db 351 DELEAVGMKHLQKLPNELSGGQQQVRSIARAFKPNLILFGDEPTGALDLEMTQIVLK 410
Qy 173 ALRALADK-GATVVVATHSPILFRESADTII 201
Db 411 QFLAIKORYKTTWIVTHNNLIQLADLVI 440

RESULT 2
Y065 MYCGE
ID Y055 MYCGE STANDARD; PRT; 466 AA.
AC P47311; Q49281;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical ABC transporter ATP-binding protein MG065.
GN MG065.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchmann J.L., Weidman J.F., Small K.V., Sandusky M., Fuhmann J.L.,
RA Nguyen D.T., Uitterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
RN [2]
RP SEQUENCE OF 393-466 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
RT sequencing.";
RL J. Bacteriol. 175:7918-7930(1993).
CC -!- SIMILARITY: Belongs to the ABC transporter family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U39686; AAC71283.1; -
DR EMBL; U02154; AAD12436.1; -
DR PIR; B64207; B64207.
DR TIGR; MG065; -
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
KW Hypothetical protein; ATP-binding; transport; Complete proteome.
FT NP_BIND 269 276 ATP (POTENTIAL).
FT CONFLICT 394 399 GDEPTG -> WWWTYV (IN REF. 2).
FT CONFLICT 394 399 GDEPTG -> WWWTYV (IN REF. 2).
SQ SEQUENCE 466 AA; 54009 MW; A585044B8A90391C CRC64;

Query Match 34.6%; Score 353; DB 1; Length 466;
Best Local Similarity 39.2%; Pred. No. 2e-22; Indels 10; Gaps 3;
Matches 73; Conservative 42; Mismatches 61;

Qy 21 SHKFLPGTMTALTGASGSGKSTLLNCLGLDLPKSSGQILVEDVLDLKLSTRKQRYKNT 80

```

```

Db 261 SHDFI-----VILGPSSGKTLINIISGMDRASSGVIVNGYNNICLNDEKLTQPKQY 315
Qy 81 VGYLFQDYALIPDRVKNLQAVEKHP-----EIPQVLHAYGLSEFEKPKVPELSSGGE 136
Db 316 VGYIFQYGLLPNLTRENIEIGANLQPDPSKRISIDALLEAVGMDSLQKLPNELSGGQ 375
Qy 137 QORTALARVLLKNPRLIILADEPTGALDNLSELVIEALRALADK-GATVVVATHSPILFRE 195
Db 376 QQRVSIARAFKPNLILFGDEPTGALDLEMTQIVLKQFLAIKORYKTTWIVTHNNLIQL 435
Qy 196 SADTII 201
Db 436 LADLVI 441

RESULT 3
MACB_ECOLI
ID MACB_ECOLI STANDARD; PRT; 648 AA.
AC P75831;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Macrolide-specific ABC-type efflux carrier.
GN MACB OR B0879.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=K12 / W3104;
RX MEDLINE=21429237; PubMed=11544226;
RA Kobayashi N., Nishino K., Yamaguchi A.;
RT "Novel macrolide-specific ABC-type efflux transporter in Escherichia
RT coli.";
RL J. Bacteriol. 183:5639-5644(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
CC -!- FUNCTION: Efflux transporter for macrolide antibiotics. Acts in
CC conjugation with macA.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC -!- SIMILARITY: Belongs to the ABC transporter family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```
DR EMBL; AB071146; BAB64542.1; -.
DR EMBL; AB000189; AAC73966.1; -.
DR EMBL; D90725; BAA35598.1; -.
DR PIR; G64826; G64826.
DR EcoGene; EG13695; macB.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR003838; DUF214.
DR Pfam; PF00005; ABC_tran; 1.
DR Pfam; PF02687; FtsX; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; transport; transmembrane; inner membrane;
KW Antibiotic resistance; Complete proteome.
FT NP_BIND 41 48
FT TRANSMEM 273 293
FT TRANSMEM 417 437
FT TRANSMEM 523 543
FT TRANSMEM 576 596
FT TRANSMEM 612 632
SQ SEQUENCE 648 AA; 70702 MW; 157B7D61ED17346 CRC64;
Query Match 34.1%; Score 347; DB 1; Length 648;
Best Local Similarity 36.9%; Pred. No. 9.4e-22;
Matches 79; Conservative 42; Mismatches 81; Indels 12; Gaps 3;
QY 1 MIEINDLKXSF-----GVRILWGLSHKFLPGTWTALTGASGSKSTLNLCLGTLDPKPS 55
DB 4 LLELKDIRRYPAGDEQEVVL-KGISLDIYAGENVAIVGASGSKSTLNLILGCLDKATS 62
QY 56 GQILVEDVLLKLTSTRQRLYKNTVGYLFQDYVALIPDRTVKFNQLAV-----EKHW 109
DB 63 GTRVACQDVATLDALAQIRHFGFIFQRVHLSHLTAEQNVFVAVYAGLERQRL 122
QY 110 PEIPQVILHVGLESFEKPFVGLSGEQRTALRVLLKNPRLADEPTGALDLNSEL 169
DB 123 LRAQELLQRLGLEDRTEVYPAQLSGGQQRVSARALMNGGVILADEPTGALDSHSEE 182
QY 170 VIEALRALADKATVTVVATHSPFRESADTIKL 203
DB 183 VMAILHOLDRGHTVIVTHDPQVAAQAEVIEI 216
RESULT 4
YXDL_BACSU STANDARD; PRT; 257 AA.
AC P42423;
DT 01-NOV-1995 (Rel. 32, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Hypothetical ABC transporter ATP-binding protein yxdl.
GN YXDL OR B65F OR BSU39640.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / BGSC141;
RX MEDLINE=95039891; PubMed=7952181;
RA Yoshida K., Sano H., Miwa Y., Ogasawara N., Fujita Y.;
RT "Cloning and nucleotide sequencing of a 15 kb region of the Bacillus
RL subtilis genome containing the iol operon.";
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / BGSC141;
RX MEDLINE=97021444; PubMed=8867804;
RA Yoshida K.-i., Fujimura M., Yanai N., Fujita Y.;
RT "Cloning and sequencing of a 23-kb region of the Bacillus subtilis
RL genome between the iol and hut operons.";
DNA Res. 2:295-301 (1995).
```

```
RN [3]
RC SEQUENCE FROM N.A.
RX STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borrius R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Borilliet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Huilo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Solido B.,
RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takenaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier P., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Darchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256 (1997).
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. STRONG, TO
CC E.COLI FTSE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL; D14399; BAA03302.1; -.
CC EMBL; D45912; BAA08315.1; -.
CC EMBL; Z59124; CAB16000.1; -.
CC PIR; A70074; A70074.
CC Subtilist; BG11128; yxdl.
CC InterPro; IPR003593; AAA_ATPase.
CC InterPro; IPR003439; ABC_transporter.
CC Pfam; PF00005; ABC_tran; 1.
CC ProDom; PD000006; ABC_transporter; 1.
CC SMART; SM00382; AAA; 1.
CC PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
CC PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
KW Hypothetical protein; ATP-binding; transport; Complete proteome.
FT NP_BIND 40 47
FT BIND 257 AA; 28722 MW; D9CEGLE9ECCIA6AD CRC64;
SQ SEQUENCE 257 AA; 33.1%; Score 337; DB 1; Length 257;
Query Match 33.1%; Score 337; DB 1; Length 257;
Best Local Similarity 37.3%; Pred. No. 2.2e-21;
Matches 76; Conservative 44; Mismatches 70; Indels 14; Gaps 5;
QY 1 MIEINDLKXSFGRVLLWGL---SHKFLPGTWTALTGASGSKSTLNLCLGTLDPKPSGQ 57
DB 4 MLEVHINKTYKGQVSYQALAKQISFSEEGFTAVMGPSGSKTTLTIITIDRPSGD 63
QY 58 ILVEDVDLLKLTSTRQRLYKNTVGYLFQDYVALIPDRTVKFNQLAVKHWKPIPOV-- 115
DB 64 ILINGENPHRLKRTKLAHFRKRLGFGVFDENLDTLTIGENIMPLTLEK--EAPSWE 121
```

FT	NP_BIND	43	50	ATP (POTENTIAL).
SQ	SEQUENCE	248 AA;	27373 MW;	711E0B8C872EB81A CRC64;
	Query Match	32.3%;	Score 329.5;	DB 1; Length 248;
	Best Local Similarity	35.5%;	Pred. No. 9.1e-21;	
	Matches	76;	Conservative	47; Mismatches 80; Indels 11; Gaps 4;
QY	1	MIENDLKSPGV----	RIILWQGLSHKFLPGTWTALTGASGSGKSTLLNCLGTLDPKSSG	56
DB	6	IVQLSNLSWTFREGETRRQVLDHITFD	EFEGFVALLGQSGSGKSTLLNCLGTLDPKSSG	65
QY	57	QILVEDVLLKSLRKORLYRKNVTGYLV	FDYALIPDRTVKFNLQL-----	AVEKHKPE 111
DB	56	DVTINGFAITQKTERDTRTLFRDQIG	IVFQFFNLPTLVLENTITLPOELAGVSQKAAV	125
QY	112	IPQ-VLHAVGLSEFPEKVPFVLSGSG	EQOORTALARKVLLKNPFIILADEPICALDLTNSLV	170
DB	126	VARDLLEKVGADRETFDPKLSGSGEQ	RVAIERALAHNPMVLADEPTGNLSDTDGDKV	185
QY	171	IEALRALA-DKGATVVVATHSP	LPRESADTIKIL 203	
DB	186	LDVLDLTRQAGKTLIVATHSP	MTQHADRNVNL 219	
RESULT 6				
Y46B	MYCGE	STANDARD;	PRT;	284 AA.
ID	Y46B MYCGE	STANDARD;	PRT;	284 AA.
AC	Q9Z570:			
DT	30-MAY-2000	(Rel. 39, Created)		
DT	30-MAY-2000	(Rel. 39, Last sequence update)		
DT	16-OCT-2001	(Rel. 40, Last annotation update)		
DE	Hypothetical ABC transporter	ATP-binding protein MG468.1.		
GN	MG468.1.			
OS	Mycoplasma genitalium.			
OC	Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.			
OX	NCBI_TaxID=2097;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 33530 / G-37;			
RX	MEDLINE=96026346; PubMed=7569993;			
RA	Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,			
RA	Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,			
RA	Peterson J.F., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,			
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,			
RA	Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,			
RA	Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;			
RT	"Whole-genome comparison of Mycobacterium tuberculosis clinical and			
RL	laboratory strains."			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CDC 1551 / Oshkosh;			
RX	MEDLINE=22206494; PubMed=12218036;			
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,			
RA	Peterson J.F., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,			
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,			
RA	Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,			
RA	Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;			
RT	"Whole-genome comparison of Mycobacterium tuberculosis clinical and			
RL	laboratory strains."			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CDC 1551 / Oshkosh;			
RX	MEDLINE=22206494; PubMed=12218036;			
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,			
RA	Peterson J.F., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,			
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,			
RA	Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,			
RA	Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;			
RT	"Whole-genome comparison of Mycobacterium tuberculosis clinical and			
RL	laboratory strains."			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=H37RV;			
RX	MEDLINE=98295987; PubMed=9634230;			
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,			
RA	Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,			
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,			
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,			
RA	Hornby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,			
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,			
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,			
RA	Sulton J.E., Taylor K., Whitehead S., Barrell B.G.;			
RT	"Deciphering the biology of Mycobacterium tuberculosis from the			
RL	complete genome sequence."			
RN	Nature 393:537-544 (1998).			
RESULT 5				
Y996	MYCTU	STANDARD;	PRT;	248 AA.
ID	Y996 MYCTU	STANDARD;	PRT;	248 AA.
AC	O53899;			
DT	16-OCT-2001	(Rel. 40, Created)		
DT	16-OCT-2001	(Rel. 40, Last sequence update)		
DT	10-OCT-2003	(Rel. 42, Last annotation update)		
DE	Hypothetical ABC transporter	ATP-binding protein Rv0986/MT1014.		
GN	Rv0986 OR MT1014 OR MV044.14.			
OS	Mycobacterium tuberculosis.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Corynebacteriineae; Mycobacteriaceae; Mycobacterium.			
OX	NCBI_TaxID=1773;</			


```
DR SMART: SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
KW Hypothetical protein; ATP-binding; Transport; Complete proteome.
FT NP_BIND 89 96 ATP (POTENTIAL).
SQ SEQUENCE 284 AA; 32256 MW; A696748EFAA3177 CRC64;

Query Match 31.9%; Score 325.5; DB 1; Length 284;
Best Local Similarity 36.3%; Pred. No. 2.3e-20;
Matches 70; Conservative 44; Mismatches 76; Indels 3; Gaps 2;

QY 14 RLWQGLSHKFLPGTMTALTGASGSKSTLLNCLGTLDPKPSGOILVEDLLKLSRKQ 73
DB 69 QLICNHINKILPGEFVILGSGSKTSLLSLSALDRPTSGDSFVCGTNTCCSDAKL 128
QY 74 RLYRKNVTGYLFQDYALIPRTVKFNQLQA--VEKHKWPPIQVILHAGVLESEEPKPVF 131
DB 129 TALRNKNVGYIFQYGLRLDLDVDDNKLALPLKRFNNLELLELLEKRRHKKVHK 188
QY 132 LSGEGQORTALRVLLKNPRIILADEPTGALDNTSELVIE-ALRALADKATVVVATHS 190
DB 189 LSGGQQQRVALARALIKEPKILFGDEPTGAVNIDISKIILQFFVEYNRDKGTIVVTEN 248
QY 191 PLFRESADTIKL 203
DB 249 EKIVELAKRVIKI 261

RESULT 7
YQIZ_BACSU STANDARD; PRT; 240 AA.
AC PS4537;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable amino-acid ABC transporter ATP-binding protein yqiz.
GN YQIZ OR BSU32960.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / JH642;
RC MEDLINE=98044033; PubMed=8969508;
RX Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M., Kobayashi Y.;
RT "Systematic sequencing of the 283 kb 210 degrees-232 degrees region of the Bacillus subtilis genome containing the skin element and many sporulation genes."
RL Microbiology 142:3103-3111(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RC MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G., Azevedo V., Bertero M.G., Besieres P., Bolotin A., Borchert S., Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S., Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M., Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A., Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T., Entian K.D., Errington J., Fabret C., Fehrlich E., Foulger D., Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N., Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G., Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A., Hilbert H., Holsappel S., Hosono S., Huilo M.F., Itaya M., Jones L., Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C., Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M., Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V., Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C., Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Ncback M., Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H., Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M., Prescecan E., Fucic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
```

Bacteria; Thermotogae; Thermotogaes; Thermotogaceae; Thermotoga.
NCBI_TaxID=2336;
[1]
SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109 / ATCC 43589;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Swinn M.L., Dodson R.J.,
Raft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
McDonald L.M., Uutterback T.R., Malek J.S., Linher C.A., Garrett M.M.,
Stewart A.L., Cottton M.D., Pratt M.S., Phillips C.A., Richardson D.,
Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RL genome sequence of Thermotoga maritima";
RI Nature 399:323-329(1999).
CC -!- SIMILARITY: Belongs to the ABC transporter family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licensese@isb-sib.ch).

DR EMBL; AE001716; AAC35439.1; --
DR PIR; H72385; H72385.
DR TIGR; TM0352; ----
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SMO0382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
KW Hypothetical protein; ATP-binding; Transport; Complete proteome.
FT NP_BIND 41 48 ATP (POTENTIAL).
SQ SEQUENCE 234 AA; 26253 MW; 8A762AA46S2BF0BD5 CRC64;

Query Match 31.7%; Score 323; DB 1; Length 234;
Best Local Similarity 36.8%; Pred.No. 3e+20;
Matches 71; Conservative 40; Mismatches 74; Indels 8; Gaps 33

DY 18 QGLSHKFLPGTMALTGASGSSTLNCLNGTLDPKSSQQLVEDVLLKLSTRKORLYR 77
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Dd 25 RGVSFEVFEGEVIITGGPSGSKSTLIHLGLCDLRPTKGVLIEGEVSRMGDRRLAQVR 84
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:

DY 78 KNTVGIFYDYALIIPRTVKFNQLAV-----EKHKWPEIQVLHAGVL-ESFEEKPVF 130
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Dd 85 NRKIGVFYSYNLLPLRALTEINVELPMIYGVPFKRRRKARELLELVGLGRLHRHP-N 143
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:

DY 131 ELSGGGQQQTALARVLLKNPRIILADEPTGALDLTNSELVIALRALADKGATVVVAATHS 190
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Dd 144 QLSSGGQQRVAIRALARANDPVFIADAEPNLTGTKEILELFRLKHGMGKTIVVTHTN 203
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:

DY 191 PLPRESADIITKL 203
: |::|
Dd 204 LEIMVDGETCIVRI 216
: |::|

RESULT 9
ID Y46B_MYCPN STANDARD; PRY; 284 AA.
AC Q50316;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical ABC transporter ATP-binding protein MG468.1 homolog
DE (X05.0f284).
GN MYCPN685 OR MP157.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI TaxID=2104;

```

106 QY 106 KHKWPEIPQVLHVAVGLESPFEE-----KPVPELSGGEQGTALARVLLKNPRILLADEPTG 160
121 Db 121 RRK-----RALECKQWAELEERFANFKP-NQSGGQQRVAARALANNPPIILADEPTG 174
161 QY 161 ALDTNSELVIEALRADLK-GATVVVATHSPFLFRESADTIKL 203
175 Db 175 ALDSKTGEKIMQLKKLNEEDGKTVVVVTHDINVARFGERRIIVL 218

RESULT 11
ID LOLD ECOLI
ID LOLD ECOLI STANDARD; PRT; 233 AA.
AC F75957; Q9R7N6;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Lipoprotein releasing system ATP-binding protein lold.
DE LOLD OR Bil17.
GN Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12.";
RT Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12.
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horuchi T.;
RA "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
RN [3]
RP SEQUENCE OF 1-11, AND CHARACTERIZATION.
RX MEDLINE=20245870; PubMed=10783239;
RA Yakushi T., Masuda K., Narita S., Matsuyama S., Tokuda H.;
RA "A new ABC transporter mediating the detachment of lipid-modified
RT proteins from membranes.";
RN Nat. Cell Biol. 2:212-218(2000).
CC -!- FUNCTION: Part of an ATP-dependent transport system responsible
CC for the release of lipoproteins targeted to the outer membrane
CC from the inner membrane. Such a release is dependent of the
CC sorting-signal (absence of an Asp at position 2 of the mature
CC lipoprotein) and of LolA.
CC -!- SUBCELLULAR LOCATION: Inner membrane-associated.
CC -!- SIMILARITY: Belongs to the ABC transporter family. Lold subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000212; AAC74201.1; ALT_INIT.
CC EMBL; D50747; BAA35937.1; ALT_INIT.
CC RSP; D58663; IG6H.
CC EcoGene; EG13440; lold.
CC InterPro; IPR003593; AAA ATPase.

```

InterPro: IPR003439; ABC_transporter.
 Pfam: PF00005; ABC_tran; 1.
 ProDom: PD000006; ABC_transporter; 1.
 SMART: SM00382; AAA; 1.
 PROSITE: PS00211; ABC_TRANSPORTER_1; 1.
 PROSITE: PS00893; ABC_TRANSPORTER_2; 1.
 NP_BIND 42 49 ATP (BY SIMILARITY).
 SEQUENCE 233 AA; 25438 MW; IC305476182FIEFD CRC64;

Query Match 30.4%; Score 309.5; DB 1; Length 233;
 Best Local Similarity 42.4%; Pred. No. 4.1e-19;
 Matches 72; Conservative 31; Mismatches 60; Indels 7; Gaps 3;

QY 27 GNTALTGASGSGKSTLLNCLGTDKSSGQILVEDVDLLKSTRKORLYRKNVTGVLFQ 86
 DB 35 GEMAIIVSSGSGKSTLLHLLGGDTPTSGDVFNGQPMKLSAAKAEARNKGLFIYQ 94
 QY 87 DYALPDPDTRVFN--LQAVKHKWPEIP---OVLVHAGVLESPEEKVPFELSGGQQRT 140
 DB 95 FPHLLPDTALENVAMPILLIGKKPAINSLAEMLKAVGLDHRANRHPSELSSGGRQV 154
 QY 141 ALARVLKXNPRIILADEPTGALDUNSELVIEALRALAD-KGATVVVVVATH 189
 DB 155 AIARALVNNPRVLADEPTGNLDARNADSIQFLLGELNQLQGTAFVLVTH 204

RESULT 12
 YF08_METUA
 ID YF08 METUA STANDARD; PRT; 224 AA.
 AC Q58903;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical ABC transporter ATP-binding protein MJ1508.
 GN MJ1508.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococcaceae; Methanocaldococcus.
 OX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weisman J.F., Sadov P.W., Hanna M.C.,
 RA Utterback T.R., Kelley J.M., Peterson J.D., Fodor P.W., Nguyen D.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii";
 RL Science 273:1058-1073 (1996).
 CC -!- SIMILARITY: Belongs to the ABC transporter family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sb-sib.ch).
 CC -----
 CC EMBL; U67591; AAB99520.1; -.
 CC DR PIR; C64488; C64488.
 CC DR TIGR; MJ1508; -.
 CC DR InterPro; IPR003593; AAA ATPase.
 CC DR InterPro; IPR003439; ABC_transporter.
 CC DR Pfam; PF00005; ABC_tran; 1.
 CC DR ProDom; PD000006; ABC_transporter; 1.
 CC DR SMART; SM00382; AAA; 1.
 CC DR

DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
 DR KW Hypothetical protein; Transport; ATP-binding; Complete proteome.
 DR FT NP_BIND 38 45 ATP (POTENTIAL).
 DR SEQUENCE 224 AA; 25263 MW; CFOB72742D24921E CRC64;

Query Match 30.3%; Score 309; DB 1; Length 224;
 Best Local Similarity 34.3%; Pred. No. 4.4e-19;
 Matches 73; Conservative 40; Mismatches 90; Indels 10; Gaps 2;

QY 1 MIEINDLKKSFG----VRILWQGLSHKFLPGTWTALTGASGSGKSTLLNCLGTLDPSSG 56
 DB 1 MIEAKNVKVIYKGEAKTIALKNLNKIEBGEFWMIPSGCGKSTLLNCLALDPTPKG 60
 QY 57 QILVEDVDLLKSTRKORLYRKNVTGVLFQDYALIPDRIVKFNQLAV-----EKHKP 110
 DB 61 EYVYKGRITSMSENERRAIFRRKISGFIQFOFHLIKTLTALENVELPMMLDERDKSYRK 120
 QY 111 EIPQVLHAGVLESPEEKVPFELSGGQQRTALARVLKXNPRIILADEPTGALDUNSELV 170
 DB 121 RAKLLEWGLGDLNHYPHQLSGGQQQORVAIRALANPKIIFADEPTGNLDSKGMV 180
 QY 171 IEALRALADKGATVVVVVATHSPLPRESADTIK 203
 DB 181 MSILKGLNKGITIIIMVTHQEQLTKVASKI 213

RESULT 13
 METN_VIBCH
 ID METN_VIBCH STANDARD; PRT; 344 AA.
 AC Q9KTJ5;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable D-methionine transport ATP-binding protein metN.
 GN METN OR VC0907.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 OX NCBI_TaxID=666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=El Tor N16961 / Serotype O1;
 RX MEDLINE=20406833; PubMed=10952301;
 RA Heideberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA Ermolova M.D., Vanathavan J., Bass S., Qin H., Dragoi I., Sellers P.,
 RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
 cholerae";
 RL Nature 406:477-483 (2000).
 CC -!- FUNCTION: Part of the binding-protein-dependent transport system
 CC metN for D-methionine. Probably responsible for energy coupling
 CC to the transport system (By similarity).
 CC -!- SUBCELLULAR LOCATION: Inner membrane-associated (Potential).
 CC -!- SIMILARITY: Belongs to the ABC transporter family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sb-sib.ch).
 CC -----
 CC EMBL; AE004174; AAF94069.1; -.
 CC DR PIR; C22866; C92266.
 CC DR TIGR; VC0907; -.
 CC DR InterPro; IPR003593; AAA_ATPase.
 CC DR InterPro; IPR003439; ABC_transporter.
 CC DR

Query Match 29.7%; Score 303; DB 1; Length 247;
 Best Local Similarity 37.7%; Pred. No. 1.6e-18;
 Matches 84; Conservative 35; Mismatches 70; Indels 34; Gaps 8;
 QY 1 MIEINDLKSGFVIRLWQGLSHKFLPGTMTALTGASGSGKSTLNCGLTLDKPSGQILV 60
 DB 1 MLTVKGLNKGFGENEILKKIDMKIEKVIAILGPSGSGKSTLNCGLTLDKPSGQILV 60
 QY 61 ED-----VDLKLSTRKQRLYRKNTVGYLFQDYALIPDRVKNLQ---LAVEK 106
 DB 61 DDFSIDFSKKVQADILKLR-----RKSPM--VFQAVHLFPHRTALENVMSGPVQVQK 111
 QY 107 HKWEIP-----QVLHVLGSLFSEKPVFELSGGQRTALARVLLKNPRIILADEPTGAL 162
 DB 112 RNKEVRKEAQLDLKVLKDKWLYFFQLSGGQQRVGARALAIQPELMFLDEPTSL 171
 QY 163 DLTNSELVIALRA---LADKATVVVATHSPULF-RESADTII 201
 DB 172 D---PELVGEVLKVIKDLANEGTMVVVTHIRKFAQEVADVI 211

RESULT 15

YXEO_BACSU STANDARD; PRT; 249 AA.
 AC PS4954;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Probable amino-acid ABC transporter ATP-binding protein yxeO.
 GN YXEO OR JP9G OR BSU39480.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 NCBI TaxID=1423;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=168 / BGSC141;
 RC MEDLINE=97021444; PubMed=8867804;
 RA Yoshida K.-I., Fujimura M., Yanai N., Fujita Y.;
 RT "Cloning and sequencing of a 23-kb region of the Bacillus subtilis
 RL genome between the *iol* and *hut* operons."
 RL DNA Res. 2:295-301(1995).
 [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=168;
 RC MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.,
 RA Gim S.-Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Giuseppe G., Guy B.J., Haga K., Haehtel J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Bianchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetlelle D., Porwollik S., Prescott A.M.,
 RA Presecan E., Pujić P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F., Soldo B.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambut R., Wedler E., Wedler H., Weitzensegger T.,
 RA Winters P., Wiput A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;

RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 RT subtilis.", 249-256(1997).
 CC FUNCTION: PROBABLY PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT
 CC SYSTEM FOR AN AMINO ACID. PROBABLY RESPONSIBLE FOR ENERGY COUPLING
 CC TO THE TRANSPORT SYSTEM.
 CC SUBCELLULAR LOCATION: Membrane-associated (Potential).
 CC SIMILARITY: Belongs to the ABC transporter family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D45912; BAA08331.1; -;
 CC EMBL; Z99124; CAB15984.1; -;
 CC PIR; A70076; A70076.
 CC Subtilist; BG11891; yxeO.
 CC InterPro; IPR003593; AAA_ATPase.
 CC InterPro; IPR003439; ABC_transporter.
 CC Pfam; PF00005; ABC_tran; 1.
 CC ProDom; PD000006; ABC_transporter; 1.
 CC SMART; SM00382; AAA; 1.
 CC PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 CC PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
 KW Hypothetical protein; ATP-binding; Transport; Membrane;
 KW Complete proteome.
 FT NP_BIND 34 41 ATP (POTENTIAL).
 SQ SEQUENCE 249 AA; 27742 MW; A63886EDE69AB80B CRC64;
 Query Match 29.7%; Score 302.5; DB 1; Length 249;
 Best Local Similarity 34.6%; Pred. No. 1.7e-18;
 Matches 74; Conservative 42; Mismatches 81; Indels 17; Gaps 5;
 QY 1 MIEINDLKSGFVIRLWQGLSHKFLPGTMTALTGASGSGKSTLNCGLTLDKPSGQILV 60
 DB 1 MLTVKIRKAFKDLVLDGLDLEVKRGEVVAITGPSGSGKSTLNCGLTLDKPSGQILV 60
 QY 61 EDVDL-LKLSTRKQRLYRKNTVGYLFQDYALIPDRVKNL--QLAVEKHKWEIP--- 113
 DB 61 GEAKLNAEKFTKZAHRLRQQTAMVFQYNLFKNKTALQNTALIVAHK---PRDEA 116
 QY 114 -----QVLHVLGSLFSEKPVFELSGGQRTALARVLLKNPRIILADEPTGALDITNSE 168
 DB 117 KRIGMEILKQVGLSHKADSYFITMSGGQQRIGIARALAVNPHAILLDEPTSLDPELVT 176
 QY 169 LVIEALRALADKATVVVATHSPULF-RESADTII 201
 DB 177 GVLQVKSIAEKQTMIIIVTHEMAFAKEVADQVI 210

Search completed: July 27, 2004, 16:35:28

Job time : 14 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 27, 2004, 16:32:35 ; Search time 38 Seconds
(without alignments)
1685.531 Million cell updates/sec

Title: US-09-868-338-9
Perfect score: 1019
Sequence: 1 MEINDLKSFVRILMOGL.....VVVATHSLFRESADTIKL 203

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	411	40.3	211	16 Q93ME0	Q93me0 clostridium
2	408	40.0	213	16 Q97RU9	Q97ru9 streptococc
3	407	39.9	212	16 Q92J8	Q92j8 clostridium
4	406	39.8	213	2 Q92HB1	Q92hb1 streptococc
5	405	39.7	213	16 Q8DQ3	Q8dqm3 streptococc
6	399	39.2	213	16 Q97NN3	Q97nn3 streptococc
7	399	39.2	213	16 Q8DND2	Q8dnd2 streptococc
8	386	37.9	213	16 Q97T44	Q97t44 streptococc
9	385	37.8	213	16 Q8DR17	Q8dr17 streptococc
10	382	37.5	207	2 Q9L550	Q9l550 lactococcus
11	380.5	37.3	210	16 Q97NR1	Q97nr1 streptococc
12	380.5	37.3	210	16 Q8DNF2	Q8dnf2 streptococc
13	375	36.8	228	16 Q9K767	Q9k767 bacillus ha
14	373	36.6	211	16 Q9CJ13	Q9cj13 lactococcus
15	371.5	36.5	248	16 Q93H50	Q93h50 streptomyc
16	366	35.9	208	16 Q9AC69	Q9ac69 staphylococ

17	361.5	35.5	226	17	O28803	O28803 archaeoglob
18	361	35.4	648	16	Q83LR7	Q83lr7 shigella fl
19	360	35.3	232	17	Q973J6	Q973j6 sulfolobus
20	358	35.1	381	16	Q8YT2	Q8yt2 brucella me
21	357.5	35.1	224	16	Q8RDY9	Q8rdy9 fusobacteri
22	356.5	35.0	246	16	Q9LOJ9	Q9loj9 streptomyc
23	356	34.9	233	16	Q8POU2	Q8pou2 streptococc
24	355	34.8	213	16	Q9KG34	Q9kg34 bacillus ha
25	354.5	34.8	228	17	O29244	O29244 archaeoglob
26	353	34.6	227	16	Q9K6Q5	Q9k6q5 bacillus ha
27	352	34.5	233	16	Q992F0	Q992f0 streptococc
28	351.5	34.5	216	16	Q8LHP0	Q8lhp0 bacillus ce
29	351	34.4	232	16	Q97IV5	Q97iv5 clostridium
30	351	34.4	233	16	Q8K7B3	Q8k7b3 streptococc
31	351	34.4	237	16	Q878P3	Q878p3 streptococc
32	350.5	34.4	300	16	Q9PR26	Q9pr26 ureaplasma
33	349.5	34.3	255	16	Q97IX2	Q97ix2 clostridium
34	349	34.2	246	17	Q9UX77	Q9ux77 sulfolobus
35	348.5	34.2	265	17	Q8PTP8	Q8ptp8 methanosarc
36	348	34.2	233	16	Q8XME3	Q8xme3 clostridium
37	348	34.2	250	16	Q8ESM1	Q8esm1 streptococc
38	348	34.2	250	16	Q8DX1	Q8dx1 streptococc
39	346	34.0	224	16	Q99XU3	Q99xu3 streptococc
40	346	34.0	231	17	O28456	O28456 archaeoglob
41	345.5	33.9	217	16	Q9PM14	Q9pm14 campylobact
42	345.5	33.9	227	16	Q894W5	Q894w5 clostridium
43	345.5	33.9	664	16	Q88UI5	Q88ui5 lactobacilli
44	344.5	33.8	227	16	Q9CNI3	Q9cni3 pasteurella
45	344	33.8	207	16	Q8Y3Q9	Q8y3q9 listeria mo

ALIGNMENTS

RESULT 1

Q93ME0	PRELIMINARY;	PRT;	211 AA.
ID	Q93ME0		
AC	Q93ME0;		
DT	01-DEC-2001 (TREMBlrel. 19, Created)		
DT	01-DEC-2001 (TREMBlrel. 19, Last sequence update)		
DT	01-OCT-2003 (TREMBlrel. 25, Last annotation update)		
DE	Probable ABC transporter.		
GN	PCP07.		
OS	Clostridium perfringens.		
OG	Plasmid pCP13.		
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;		
OC	Clostridium.		
OX	NCBI_taxid=1502;		
RP	[1]		
RN	SEQUENCE FROM N.A.		
RC	STRAIN=13 / Type A;		
RX	MEDLINE=21664373; PubMed=11792842;		
RA	Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,		
RA	Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;		
RT	Complete genome sequence of Clostridium perfringens, an anaerobic		
RT	flesh-eater.;		
RL	Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).		
DR	EMBL; AP003515; BAB62445.1;		
DR	GO; GO:0046821; C:extrachromosomal DNA; IEA.		
DR	GO; GO:0016020; C:membrane; IEA.		
DR	GO; GO:0005524; F:ATP binding; IEA.		
DR	GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.		
DR	GO; GO:0000166; F:nucleotide binding; IEA.		
DR	GO; GO:0006810; P:transport; IEA.		
DR	InterPro; IPR003593; AAA ATPase.		
DR	InterPro; IPR003439; ABC transporter.		
DR	Pfam; PF00005; ABC tran; 1.		
DR	ProDom; PD000006; ABC transporter; 1.		
DR	SMART; SM00382; AAA; 1.		
DR	PROSITE; PS00211; ABC TRANSPORTER_1; 1.		
DR	PROSITE; PS00893; ABC TRANSPORTER_2; 1.		
KW	Plasmid; Complete proteome.		
SQ	SEQUENCE 211 AA; 23971 MW; D490252EB93B5724 CRC64;		


```
Q9ZHB1
ID Q9ZHB1 PRELIMINARY; PRT; 213 AA.
AC Q9ZHB1;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RC STRAIN=GS4;
RC MEDLINE=99061199; PubMed=9846742;
RX Massidda O., Anderluzzi D., Friedli L., Feger G.;
RA "unconventional organization of the division and cell wall gene
RT cluster of Streptococcus pneumoniae.";
RL Microbiology 144:3069-3078(1998).
CC -I- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL; AF068902; AAC95448.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:000524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.
DR GO; GO:000166; P:nucleotide binding; IEA.
DR GO; GO:0006810; P:transporter; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
KW Hypothetical protein; ATP-binding; Transporter.
SQ SEQUENCE 213 AA; 23962 MW; 2939F15CB0044F4 CRC64;

Query Match 39.8%; Score 406; DB 2; Length 213;
Best Local Similarity 41.9%; Pred. No. 2,le-25;
Matches 88; Conservative 43; Mismatches 69; Indels 10; Gaps 3;

QY 1 MIEINDLKSPGVRIILWQSLSHKPLFGTWTALTGASGSKTLNCLGTLDPKSSGQILV 60
DB 1 MIELKNISKFGSRQLFSDTNLHFEGGKIYALIGTSCGKTTLLNMGSLPEYDKGIIY 60
QY 61 EDVDLLKSTRKQRLYKNTVGYLFQDYALIPDRTVKFNQLAV-----EKHKWPEIPQ 114
DB 61 DGTSLKDI---KSSVFRDYGLYFQDFGLIESQTVKENLNGLVGKKLKEKISLMKQ 117
QY 115 VLHVGLESPE-EKPVFELSGGEGQRTALARVLLKNPRIILADEPTGALDLTNSLVIEA 173
DB 118 ALNRVNLVYLDLKPQIFELSGGGAQRVALAKIILKDPPIILADEPTASLDPKNSSELLSI 177
QY 174 LRALAKGATVVVATHSPFPRESADTIKL 203
DB 178 LESLKNPNRTIIATHNPLNWEQDVQVIRV 207

RESULT 5
Q8DQM3
ID Q8DQM3 PRELIMINARY; PRT; 213 AA.
AC Q8DQM3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE ABC transporter ATP-binding protein-sodium transport.
GN ABC-NBD OR SFR0602.
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RC STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tetzelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
```

```
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Gerlinger C.,
RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA Leblanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAhren S.M., McHenry M., McLeaster K., Mundy C.W., Niclas T.I.,
RA Norris F.H., O'Garra M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rosteck P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717(2001).
DR EMBL; AE008439; AAX99406.1; -.
DR FIR; B97947; B97947.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:000524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.
DR GO; GO:000166; P:nucleotide binding; IEA.
DR GO; GO:0006810; P:transporter; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Complete proteome.
SQ SEQUENCE 213 AA; 23972 MW; C4989CCCB0062965 CRC64;

Query Match 39.7%; Score 405; DB 16; Length 213;
Best Local Similarity 41.9%; Pred. No. 2,6e-25;
Matches 88; Conservative 43; Mismatches 69; Indels 10; Gaps 3;

QY 1 MIEINDLKSPGVRIILWQSLSHKPLFGTWTALTGASGSKTLNCLGTLDPKSSGQILV 60
DB 1 MIELKNISKFGSRQLFSDTNLHFEGGKIYALIGTSCGKTTLLNMGSLPEYDKGIIY 60
QY 61 EDVDLLKSTRKQRLYKNTVGYLFQDYALIPDRTVKFNQLAV-----EKHKWPEIPQ 114
DB 61 DGTSLKDI---KPSVFRDYGLYFQDFGLIESQTVKENLNGLVGKKLKEKISLMKQ 117
QY 115 VLHVGLESPE-EKPVFELSGGEGQRTALARVLLKNPRIILADEPTGALDLTNSLVIEA 173
DB 118 ALNRVNLVYLDLKPQIFELSGGGAQRVALAKIILKDPPIILADEPTASLDPKNSSELLSI 177
QY 174 LRALAKGATVVVATHSPFPRESADTIKL 203
DB 178 LESLKNPNRTIIATHNPLNWEQDVQVIRV 207

RESULT 6
Q97NN3
ID Q97NN3 PRELIMINARY; PRT; 213 AA.
AC Q97NN3;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE ABC transporter, ATP-binding protein.
GN SPI987.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RC STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tetzelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
```

RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
 RT "Complete genome sequence of a virulent isolate of Streptococcus
 RT pneumoniae".
 RL Science 293:498-506(2001).
 DR EMBL; AE007488; AAK76054.1; -.
 DR PIR; E95232; E95232.
 DR PIR; G98096; G98096.
 DR TIGR; SP1987; -.
 DR GO; GO:0016020; C-membrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR003439; ABC_transporter.
 DR Pfam; PF00005; ABC_tran; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
 DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
 KW ATP-binding; Complete proteome.
 SQ SEQUENCE 213 AA; 23920 MW; F6BE2EA4EB87FDB7 CRC64;
 Query Match 39.2%; Score 399; DB 16; Length 213;
 Best Local Similarity 42.9%; Pred. No. 7.9e-25;
 Matches 90; Conservative 40; Mismatches 70; Indels 10; Gaps 3;
 QY 1 MIEINDLKSGFVRLWQGLSHKFLPGTMTALTGASGSGKSTLLNCLGTLDPKSSGQILV 60
 Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
 1 MIDIQGLEKFNDRIFAISGLNKLKQKVALIGKSGKSTLLNCLGTLDPKSSGQILV 60
 QY 61 EDVLLKLSRKQRLYKNTVGYLFQDYALIPDRVKNLQLAVERKHWPPIQVLEHVG 120
 Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
 61 QGKDLKTIPTRE---YFDDQMGYLFQNGLENSQIKENLDLGFVGQKISKVERLERQVG 117
 QY 121 -LESFE-----EKPVFELSGGEOQTALARVLKPNRIILADEPTGALDITNSSELV 173
 Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
 118 ALEKVNLYGLDLEQKIYTLSGGEAQRVALAKTILKNPPLILADEPTAALDPENSEEVMNL 177
 QY 174 LRALADKGTAVVATHSPFPRESADTIKL 203
 Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
 178 LVLDKDNRIIIATHNPLVWKNKADEIIDM 207
 RESULT 7
 Q8DND2 PRELIMINARY; PRT; 213 AA.
 ID Q8DND2
 AC Q8DND2
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE ABC transporter ATP-binding protein-unknown substrate.
 GN ABC-NBD OR SP1801.
 OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=171101;
 RN [1]
 SEQUENCE FROM N.A.
 RC MEDLINE=21429245; PubMed=11544234;
 RX Hoskins J., Albom W.E. Jr., Arnold J., Blaszcak L.C., Bargett S.,
 RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller R.C., Geringer C.,
 RA Gilmour R., Glass J.S., Knoch H., Kraft A.R., Lagace R.E.,
 RA LeBlanc D.J., Lee L.N., Leikowitz E.J., Lu J., Matsushima P.,
 RA McAhren S.M., McHenry M., McKeen K., Mundy C.W., Niclas T.I.,
 RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,
 RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
 RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostock P.R. Jr., Skatrud P.L.,
 RA Glass J.L.;
 RT "Genome of the bacterium Streptococcus pneumoniae strain R6";
 RL J. Bacteriol. 183:5709-5717(2001).
 DR EMBL; AE008544; AAL00604.1; -.
 DR PIR; E95232; E95232.
 DR PIR; G98096; G98096.
 DR GO; GO:0016020; C-membrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.

DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
 DR GO; GO:000166; P:nucleotide binding; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR003593; AAA_AIPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
 KW ATP-binding; Complete proteome.
 SQ SEQUENCE 213 AA; 23920 MW; F6BE2EA4EB87FDB7 CRC64;
 Query Match 39.2%; Score 399; DB 16; Length 213;
 Best Local Similarity 42.9%; Pred. No. 7.9e-25;
 Matches 90; Conservative 40; Mismatches 70; Indels 10; Gaps 3;
 QY 1 MIEINDLKSGFVRLWQGLSHKFLPGTMTALTGASGSGKSTLLNCLGTLDPKSSGQILV 60
 Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
 1 MIDIQGLEKFNDRIFAISGLNKLKQKVALIGKSGKSTLLNCLGTLDPKSSGQILV 60
 QY 61 EDVLLKLSRKQRLYKNTVGYLFQDYALIPDRVKNLQLAVERKHWPPIQVLEHVG 120
 Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
 61 QGKDLKTIPTRE---YFDDQMGYLFQNGLENSQIKENLDLGFVGQKISKVERLERQVG 117
 QY 121 -LESFE-----EKPVFELSGGEOQTALARVLKPNRIILADEPTGALDITNSSELV 173
 Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
 118 ALEKVNLYGLDLEQKIYTLSGGEAQRVALAKTILKNPPLILADEPTAALDPENSEEVMNL 177
 QY 174 LRALADKGTAVVATHSPFPRESADTIKL 203
 Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
 178 LVLDKDNRIIIATHNPLVWKNKADEIIDM 207
 RESULT 8
 Q97T44 PRELIMINARY; PRT; 213 AA.
 ID Q97T44
 AC Q97T44
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Amino acid ABC transporter, ATP-binding protein, putative.
 GN SP0111.
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1313;
 RN [1]
 SEQUENCE FROM N.A.
 RC STRAIN=ATCC BAA-334 / TIGR4;
 RX MEDLINE=21357209; PubMed=11463916;
 RA Tetelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
 RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
 RA Holtzapple E., Khouri H., Wolf A.M., Uterback T.R., Hansen C.L.,
 RA McDonald L.A., Kholm T.V., Angiuoli S., Dickinson T., Hickey E.K.,
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
 RT "Complete genome sequence of a virulent isolate of Streptococcus
 RT pneumoniae";
 RL Science 293:498-506(2001).
 DR EMBL; AE007327; AAK74298.1; -.
 DR PIR; A95013; A95013.
 DR TIGR; SP0111; -.
 DR GO; GO:0016020; C-membrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
 DR GO; GO:000166; P:nucleotide binding; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR003593; AAA_AIPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR Pfam; PF00005; ABC_tran; 1.

DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Complete proteome.
SQ SEQUENCE 213 AA; 23640 MW; E35696AEB3677399 CRC64;

Query Match 37.9%; Score 386; DB 16; Length 213;
Best Local Similarity 40.2%; Pred. No. 9.2e-24;
Matches 84; Conservative 45; Mismatches 74; Indels 6; Gaps 1;

QY 1 MIEINDLKSGFVRILWQGLSHKFLPGTMTALTGASGSGKSTLLNCLGTLDPKSSGQILV 60
DB 1 MIELKNITKIGGVKVLNLSLRIDQDLVAIVGKSGSKTLLNLLGLIDGYSRVEI 60
QY 61 EDVDLLKSTRKQRLYRKNVTGYLFDYALIPDRTVKFNQLAV-----EKHKWPEIQ 114
DB 61 FGQTNLAVNSAKSQTIIEHISYLFQNFALIDDETVEYNLMALKYVKLPKDKKLKVEE 120
QY 115 VLHVGLESFEEKPVPFELSGEQOQTALARVLLKPNRIILADEPTGALDTNSELVIEAL 174
DB 121 ILERVGLSATLHQRVSELSGGEQOQIAVARAILKPSQILADEPTGSLDPENRDLVKFL 180
QY 175 RALADKAGTAVVATHSPFLFRESADTIKL 203
DB 181 LEMNREGKTVIIVTHDAYVAQOCHRIEL 209

RESULT 9
QY 175 RALADKAGTAVVATHSPFLFRESADTIKL 203
DB 181 LEMNREGKTVIIVTHDAYVAQOCHRIEL 209

DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Complete proteome.
SQ SEQUENCE 213 AA; 23640 MW; E35696AEB3677399 CRC64;

Query Match 37.9%; Score 386; DB 16; Length 213;
Best Local Similarity 40.2%; Pred. No. 9.2e-24;
Matches 84; Conservative 45; Mismatches 74; Indels 6; Gaps 1;

QY 1 MIEINDLKSGFVRILWQGLSHKFLPGTMTALTGASGSGKSTLLNCLGTLDPKSSGQILV 60
DB 1 MIELKNITKIGGVKVLNLSLRIDQDLVAIVGKSGSKTLLNLLGLIDGYSRVEI 60
QY 61 EDVDLLKSTRKQRLYRKNVTGYLFDYALIPDRTVKFNQLAV-----EKHKWPEIQ 114
DB 61 FGQTNLAVNSAKSQTIIEHISYLFQNFALIDDETVEYNLMALKYVKLPKDKKLKVEE 120
QY 115 VLHVGLESFEEKPVPFELSGEQOQTALARVLLKPNRIILADEPTGALDTNSELVIEAL 174
DB 121 ILERVGLSATLHQRVSELSGGEQOQIAVARAILKPSQILADEPTGSLDPENRDLVKFL 180
QY 175 RALADKAGTAVVATHSPFLFRESADTIKL 203
DB 181 LEMNREGKTVIIVTHDAYVAQOCHRIEL 209

RESULT 9
QY 175 RALADKAGTAVVATHSPFLFRESADTIKL 203
DB 181 LEMNREGKTVIIVTHDAYVAQOCHRIEL 209

DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Complete proteome.
SQ SEQUENCE 213 AA; 23640 MW; E35696AEB3677399 CRC64;

Query Match 37.9%; Score 386; DB 16; Length 213;
Best Local Similarity 40.2%; Pred. No. 9.2e-24;
Matches 84; Conservative 45; Mismatches 74; Indels 6; Gaps 1;

QY 1 MIEINDLKSGFVRILWQGLSHKFLPGTMTALTGASGSGKSTLLNCLGTLDPKSSGQILV 60
DB 1 MIELKNITKIGGVKVLNLSLRIDQDLVAIVGKSGSKTLLNLLGLIDGYSRVEI 60
QY 61 EDVDLLKSTRKQRLYRKNVTGYLFDYALIPDRTVKFNQLAV-----EKHKWPEIQ 114
DB 61 FGQTNLAVNSAKSQTIIEHISYLFQNFALIDDETVEYNLMALKYVKLPKDKKLKVEE 120
QY 115 VLHVGLESFEEKPVPFELSGEQOQTALARVLLKPNRIILADEPTGALDTNSELVIEAL 174
DB 121 ILERVGLSATLHQRVSELSGGEQOQIAVARAILKPSQILADEPTGSLDPENRDLVKFL 180
QY 175 RALADKAGTAVVATHSPFLFRESADTIKL 203
DB 181 LEMNREGKTVIIVTHDAYVAQOCHRIEL 209

RESULT 9
QY 175 RALADKAGTAVVATHSPFLFRESADTIKL 203
DB 181 LEMNREGKTVIIVTHDAYVAQOCHRIEL 209

DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Complete proteome.
SQ SEQUENCE 213 AA; 23640 MW; E35696AEB3677399 CRC64;

Query Match 37.9%; Score 386; DB 16; Length 213;
Best Local Similarity 40.2%; Pred. No. 9.2e-24;
Matches 84; Conservative 45; Mismatches 74; Indels 6; Gaps 1;

QY 1 MIEINDLKSGFVRILWQGLSHKFLPGTMTALTGASGSGKSTLLNCLGTLDPKSSGQILV 60
DB 1 MIELKNITKIGGVKVLNLSLRIDQDLVAIVGKSGSKTLLNLLGLIDGYSRVEI 60
QY 61 EDVDLLKSTRKQRLYRKNVTGYLFDYALIPDRTVKFNQLAV-----EKHKWPEIQ 114
DB 61 FGQTNLAVNSAKSQTIIEHISYLFQNFALIDDETVEYNLMALKYVKLPKDKKLKVEE 120
QY 115 VLHVGLESFEEKPVPFELSGEQOQTALARVLLKPNRIILADEPTGALDTNSELVIEAL 174
DB 121 ILERVGLSATLHQRVSELSGGEQOQIAVARAILKPSQILADEPTGSLDPENRDLVKFL 180
QY 175 RALADKAGTAVVATHSPFLFRESADTIKL 203
DB 181 LEMNREGKTVIIVTHDAYVAQOCHRIEL 209

RESULT 9
QY 175 RALADKAGTAVVATHSPFLFRESADTIKL 203
DB 181 LEMNREGKTVIIVTHDAYVAQOCHRIEL 209

DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Complete proteome.
SQ SEQUENCE 213 AA; 23640 MW; E35696AEB3677399 CRC64;

Query Match 37.9%; Score 386; DB 16; Length 213;
Best Local Similarity 40.2%; Pred. No. 9.2e-24;
Matches 84; Conservative 45; Mismatches 74; Indels 6; Gaps 1;

QY 1 MIEINDLKSGFVRILWQGLSHKFLPGTMTALTGASGSGKSTLLNCLGTLDPKSSGQILV 60
DB 1 MIELKNITKIGGVKVLNLSLRIDQDLVAIVGKSGSKTLLNLLGLIDGYSRVEI 60
QY 61 EDVDLLKSTRKQRLYRKNVTGYLFDYALIPDRTVKFNQLAV-----EKHKWPEIQ 114
DB 61 FGQTNLAVNSAKSQTIIEHISYLFQNFALIDDETVEYNLMALKYVKLPKDKKLKVEE 120
QY 115 VLHVGLESFEEKPVPFELSGEQOQTALARVLLKPNRIILADEPTGALDTNSELVIEAL 174
DB 121 ILERVGLSATLHQRVSELSGGEQOQIAVARAILKPSQILADEPTGSLDPENRDLVKFL 180
QY 175 RALADKAGTAVVATHSPFLFRESADTIKL 203
DB 181 LEMNREGKTVIIVTHDAYVAQOCHRIEL 209

RESULT 9
QY 175 RALADKAGTAVVATHSPFLFRESADTIKL 203
DB 181 LEMNREGKTVIIVTHDAYVAQOCHRIEL 209

DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Complete proteome.
SQ SEQUENCE 213 AA; 23640 MW; E35696AEB3677399 CRC64;

Query Match 37.9%; Score 386; DB 16; Length 213;
Best Local Similarity 40.2%; Pred. No. 9.2e-24;
Matches 84; Conservative 45; Mismatches 74; Indels 6; Gaps 1;

QY 1 MIEINDLKSGFVRILWQGLSHKFLPGTMTALTGASGSGKSTLLNCLGTLDPKSSGQILV 60
DB 1 MIELKNITKIGGVKVLNLSLRIDQDLVAIVGKSGSKTLLNLLGLIDGYSRVEI 60
QY 61 EDVDLLKSTRKQRLYRKNVTGYLFDYALIPDRTVKFNQLAV-----EKHKWPEIQ 114
DB 61 FGQTNLAVNSAKSQTIIEHISYLFQNFALIDDETVEYNLMALKYVKLPKDKKLKVEE 120
QY 115 VLHVGLESFEEKPVPFELSGEQOQTALARVLLKPNRIILADEPTGALDTNSELVIEAL 174
DB 121 ILERVGLSATLHQRVSELSGGEQOQIAVARAILKPSQILADEPTGSLDPENRDLVKFL 180
QY 175 RALADKAGTAVVATHSPFLFRESADTIKL 203
DB 181 LEMNREGKTVIIVTHDAYVAQOCHRIEL 209

RESULT 9
QY 175 RALADKAGTAVVATHSPFLFRESADTIKL 203
DB 181 LEMNREGKTVIIVTHDAYVAQOCHRIEL 209

DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Complete proteome.
SQ SEQUENCE 213 AA; 23640 MW; E35696AEB3677399 CRC64;

Query Match 37.9%; Score 386; DB 16; Length 213;
Best Local Similarity 40.2%; Pred. No. 9.2e-24;
Matches 84; Conservative 45; Mismatches 74; Indels 6; Gaps 1;

QY 1 MIEINDLKSGFVRILWQGLSHKFLPGTMTALTGASGSGKSTLLNCLGTLDPKSSGQILV 60
DB 1 MIELKNITKIGGVKVLNLSLRIDQDLVAIVGKSGSKTLLNLLGLIDGYSRVEI 60
QY 61 EDVDLLKSTRKQRLYRKNVTGYLFDYALIPDRTVKFNQLAV-----EKHKWPEIQ 114
DB 61 FGQTNLAVNSAKSQTIIEHISYLFQNFALIDDETVEYNLMALKYVKLPKDKKLKVEE 120

QY 174 LRALADKATVVVATHSPFRESADTIKL 203
 Db 178 LLSLVKENTVIIATHSHVNRVDSIINL 207

RESULT 11

Q37NRI
 ID Q37NRI PRELIMINARY; PRT; 210 AA.
 AC Q37NRI
 DT 01-OCT-2001 (TREMELrel. 18, Created)
 DT 01-OCT-2001 (TREMELrel. 18, Last sequence update)
 DE ABC transporter, ATP-binding protein.
 GN SP1957.
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1313;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC BAA-334 / TIGR4;
 RX MEDLINE=21357209; PubMed=11463916;
 RA Tetelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
 RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
 RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
 RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
 RT "Complete genome sequence of a virulent isolate of Streptococcus pneumoniae";
 RL Science 293:498-506(2001).
 DR EMEL; AE007485; AAK76024.1; --
 DR PIR; C98093; C98093.
 DR TIGR; SP1957; --
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0008170; F:N-methyltransferase activity; IEA.
 DR GO; GO:0006306; P:DNA methylation; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR003439; ABC transporter.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
 DR PROSITE; PS00092; N6_MTASE; 1.
 KW ATP-binding; Complete proteome.
 SQ SEQUENCE 210 AA; 23751 MW; EC0373F9B66FE167 CRC64;

Query Match 37.3%; Score 380.5; DB 16; Length 210;
 Best Local Similarity 41.0%; Pred. No. 2.5e-23;
 Matches 86; Conservative 41; Mismatches 72; Indels 11; Gaps 4;
 QY 1 MIEINDLKSGVIRILMOGLSHKFLPTMTALTGASGSGKSTLNCIGTLDPKSSGQILV 60
 Db 1 MIELQVSKSGFERELFNSMTFAGKVYALIGSSGSGKTLNMMIGKLE-PYDGTIFY 59
 QY 61 EDVLLKLSRKQRLYKNTVGYLFQDYALIPDRVTNFNLQAVEKKWPEI-----PQ 114
 Db 60 RGKD---LANYKSSDFRHEGLYLFQNGFLIENQSIENKLGIGOKLSRSEQLRQKQ 116
 QY 115 VLHVGLESFE-EKPVFELSGEQORTALARVLKNPRIILADEPTGALDITNSLVEIA 173
 Db 117 ALEQVGLVYLLDKRIIFELSGESQORVALAKIILKNPFIILADEPTASIDPATSQLIMEI 176
 QY 174 LRALADKATVVVATHSPFRESADTIKL 203

Db 177 LLSLRDDNRLIIITHNPAINWMADEVFTM 206

RESULT 12

Q8DNF2
 ID Q8DNF2 PRELIMINARY; PRT; 210 AA.
 AC Q8DNF2
 DT 01-MAR-2003 (TREMELrel. 23, Created)
 DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
 DE ABC transporter ATP-binding protein-unknown substrate.
 GN ABC-NBD OR SP1773.
 OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=171101;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21429245; PubMed=11544234;
 RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
 RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
 RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
 RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
 RA McAnen S.M., McHenry M., Mcleaster K., Mundy C.W., Niclas T.I.,
 RA Norris F.H., O'Gara M.E., Peery R.B., Robertson G.T., Rockey P.,
 RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
 RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,
 RA Glass J.I.;
 RT "Genome of the bacterium Streptococcus pneumoniae strain R6";
 RL J. Bacteriol. 183:5709-5717(2001).
 DR EMEL; AE008541; AAL00576.1; --
 DR PIR; C98093; C98093.
 DR PIR; G95228; G95228.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0008170; F:N-methyltransferase activity; IEA.
 DR GO; GO:0006306; P:DNA methylation; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR003593; AAA ATPase.
 DR InterPro; IPR003439; ABC transporter.
 DR InterPro; IPR002052; N6_Mtase.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
 DR PROSITE; PS00092; N6_MTASE; 1.
 KW ATP-binding; Complete proteome.
 SQ SEQUENCE 210 AA; 23751 MW; EC0373F9B66FE167 CRC64;

Query Match 37.3%; Score 380.5; DB 16; Length 210;
 Best Local Similarity 41.0%; Pred. No. 2.5e-23;
 Matches 86; Conservative 41; Mismatches 72; Indels 11; Gaps 4;
 QY 1 MIEINDLKSGVIRILMOGLSHKFLPTMTALTGASGSGKSTLNCIGTLDPKSSGQILV 60
 Db 1 MIELQVSKSGFERELFNSMTFAGKVYALIGSSGSGKTLNMMIGKLE-PYDGTIFY 59
 QY 61 EDVLLKLSRKQRLYKNTVGYLFQDYALIPDRVTNFNLQAVEKKWPEI-----PQ 114
 Db 60 RGKD---LANYKSSDFRHEGLYLFQNGFLIENQSIENKLGIGOKLSRSEQLRQKQ 116
 QY 115 VLHVGLESFE-EKPVFELSGEQORTALARVLKNPRIILADEPTGALDITNSLVEIA 173
 Db 117 ALEQVGLVYLLDKRIIFELSGESQORVALAKIILKNPFIILADEPTASIDPATSQLIMEI 176
 QY 174 LRALADKATVVVATHSPFRESADTIKL 203
 Db 177 LLSLRDDNRLIIITHNPAINWMADEVFTM 206


```

RC STRAIN=WA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RT Sakaki Y., Hattori M., Omura S.;
RR "Complete genome sequence and comparative analysis of the industrial
RL microorganism Streptomyces avermitilis.";
RT Nat. Biotechnol. 21:526-531(2003)
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR ENBL; AB070951; BAB69341.1; -.
DR DR ENBL; AP005035; BAC71346.1; -.
DR GO; GO:0016020; C.membrane; IEA.
DR GO; GO:0005524; F.ATP binding; IEA.
DR GO; GO:0004009; F.ATP-binding cassette (ABC) transporter acti...; IEA.
DR GO; GO:000166; F.nucleotide binding; IEA.
DR GO; GO:0006810; P.transport; IEA.
DR InterPro; IPR003593; AAA_ATase.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS0893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Transport; Complete Proteome.
SQ SEQUENCE 248 AA; 46276 MW; 43F62AE02BB357E9 CRC64;

Query Match          36.5%; Score 371.5; DB 16; Length 248;
Best Local Similarity 41.0%; Pred. No. 1.7e-22;
Matches      87; Conservative 37; Mismatches 77; Indels 11; Gaps 4

QY    2  IEINDLKSEFG-----VRILWGLSHKPLPGMTALTGASGGSKTLNCLGTLDKPSSG 56
Db     9  IQRSVSRIRYGAGGNVTALDQ-VSLAPRGFTTAVMGPSGSGSKTLQCAGLDRPTSG 67
QY   57  QILVEDVDLLKLSTRQRLYRKNTVGYLFQDYALIPDRIVRPN----LQLAVEKHKWEI 111
Db    68  SVTVGDTETLTLSLTKLTLLRRDRIGCFVFAENILLPSLTAEQNVALPURLAGRPRKTEV 174
QY   113  POVLHVAVGLESFEEXPEVPFELSGGEORTALARVLTKNPRITILADEPTCALDLTSELVIE 174
Db   128  REVLAQVGLGDRAGRHPTEMSCGQORVALARALITRDVLPFGDEPTGALDSQTSEVIT 181
QY   173  ALRALAD-KGATVVVATHSPLEPRESADIITKL 203
Db   188  LLRGMWDSEGQTIVMTHDPVAASVADRNVFL 219

```

Search completed: July 27, 2004, 16:36:25
Job time : 47 secs